PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 4 : C12P 21/00, C12N 15/00, 5/00 C07H 15/12, A23J 1/00 A61K 39/40

A1

(11) International Publication Number: WO 86/ 02100

(43) International Publication Date: 10 April 1986 (10.04.86)

(21) International Application Number: PCT/US85/01900

(22) International Filing Date: 1 October 1985 (01.10.85)

(31) Priority Application Number: 656,590

(32) Priority Date: 1 October 1984 (01.10.84)

(33) Priority Country: US

(71) Applicants: SANDOZ LTD. [CH/CH]; Lichtstrasse 35, CH-4002 Basel (CH). THE REGENTS OF THE UNIVERSITY OF CALIFORNIA [US/US]; 2199 Addison Street, Berkeley, CA 94720 (US).

(72) Inventors: BROWN, Eugene ; 1388 Walnut Street, Newton Highlands, MA 02161 (US). CLARK, Steven, C. ; 122 Johnson Road, Winchester, MA 01890 (US). GOLDE, David, W. ; 10718 Lindbrook Avenue, Los Angeles, CA 90024 (US). CASSON, Judith, C. ; 631 N. Cherokee Avenue, Los Angeles, CA 90004 (US). HEWICK, Rodney ; 16 Woodcliffe Road, Lexington, MA 02173 (US). WESTBROOK, Carol, A.; 5413 Ridgewood Court, Chicago, IL 60615 (US).

(74) Agents: SHARKIN, Gerald, D. et al.; Sandoz, Inc., 59 Route 10, East Hanover, NJ 07936 (US).

(81) Designated States: AT (European patent), BE (European patent), CH (European patent), DE (European patent), FR (European patent), GB (European patent), IT (European patent), JP, LU (European patent), NL (European patent), SE (European patent).

Published
With international search report.

(54) Title: VECTORS CONTAINING A GENE FOR A PROTEIN HAVING ERYTHROID-POTENTIATING ACTIVITY AND RECOMBINANT DNA METHODS FOR PRODUCING SAID PROTEIN

(57) Abstract

A protein having erythroid potentiating activity (EPA). The EPA protein has a biological activity of at least about 10^6 units per mg of protein and has an apparent molecular weight of about 28,000 daltons as determined by SDS-PAGE analysis on 10 percent polyacrylamide gel. Also described is a cloned EPA gene coding for such EPA protein, a recombinant vector containing the EPA gene, a microorganism or mammalian cell transformed with such a recombinant vector, and a method for producing EPA by expressing said EPA gene by culturing said microorganism or mammalian cell. Purified EPA stimulates the growth of both early and late erythroid precursors from human bone marrow, as well as colony formation by the K562 erythroleukemia cell line.

```
                10                  20                  30                  40
CCCGCAGATC CAGCGCCCAG AGAGACACCA GAGAACCCAC C ATG GCC CCC TTT
                                                      MET Ala Pro Phe

 56                  71                  86
GAG CCC CTG GCT TCT GGC ATC CTG TTG TTG CTG TGG CTG ATA GCC
Glu Pro Leu Ala Ser Gly Ile Leu Leu Leu Leu Trp Leu Ile Ala

101                 116                 131
CCC AGC AGG GCC TGC ACG TGT GTC CCA CCC CAC CCA CAG ACG GCC
Pro Ser Arg Ala Cys Thr Cys Val Pro Pro His Pro Gln Thr Ala

146                 161                 176
TTC TGC AAT TCC GAC CTC GTC ATC AGG GCC AAG TTC GTG GGG ACA
Phe Cys Asn Ser Asp Leu Val Ile Arg Ala Lys Phe Val Gly Thr

191                 206                 221
CCA GAA GTC AAC CAG ACC ACC TTA TAC CAG CGT TAT GAG ATC AAG
Pro Glu Val Asn Gln Thr Thr Leu Tyr Gln Arg Tyr Glu Ile Lys

236                 251                 266
ATG ACC AAG ATG TAT AAA GGG TTC CAA GCC TTA GGG GAT GCC GCT
MET Thr Lys MET Tyr Lys Gly Phe Gln Ala Leu Gly Asp Ala Ala

281                 296                 311
GAC ATC CGG TTC GTC TAC ACC CCC GCC ATG GAG AGT GTC TGC GGA
Asp Ile Arg Phe Val Tyr Thr Pro Ala MET Glu Ser Val Cys Gly

326                 341                 356
TAC TTC CAC AGG TCC CAC AAC CGG AGC GAG GAG TTT CTC ATT GCT
Tyr Phe His Arg Ser His Asn Arg Ser Glu Glu Phe Leu Ile Ala

371                 386                 401
GGA AAA CTG CAG GAT GGA CTC TTG CAC ATC ACT ACC TGC AGT TTT
Gly Lys Leu Gln Asp Gly Leu Leu His Ile Thr Thr Cys Ser Phe

416                 431                 446
GTG GCT CCC TGG AAC AGC CTG AGC TTA GCT CAG CGC CGG GGC TTC
Val Ala Pro Trp Asn Ser Leu Ser Leu Ala Gln Arg Arg Gly Phe

461                 476                 491
ACC AAG ACC TAC ACT GTT GGC TGT GAG GAA TGC ACA GTG TTT CCC
Thr Lys Thr Tyr Thr Val Gly Cys Glu Glu Csy Thr Val Phe Pro

506                 521                 536
TGT TTA TCC ATC CCC TGC AAA CTG CAG AGT GGC ACT CAT TGC TTG
Cys Leu Ser Ile Pro Cys Lys Leu Gln Ser Gly Thr His Cys Leu

551                 566                 581
TGG ACG GAC CAG CTC CTC CAA GGC TCT GAA AAG GGC TTC CAG TCC
Trp Thr Asp Gln Leu Leu Gln Gly Ser Glu Lys Gly Phe Glu Ser

596                 611                 626
CGT CAC CTT GCC TGC CTG CCT CGG GAG CCA GGG CTG TGC ACC TGG
Arg His Leu Ala Cys Leu Pro Arg Glu Pro Gly Leu Cys Thr Trp

641                 656                 675         685
CAG TCC CTG CGG TCC CAG ATA GCC TGA ATCCTGCCCG GAGTGGAAGC
Gln Ser Leu Arg Ser Gln Ile Ala

      695         705         715         725         735
TGAAGCCTGC ACAGTGTCCA CCCTGTTCCC ACTCCCATCT TTCTTCCGGA

          745         755         765
CAATGAAATA AAGAGTTACA CCAGCAAAAA AAAAA
```

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GA	Gabon	MR	Mauritania
AU	Australia	GB	United Kingdom	MW	Malawi
BB	Barbados	HU	Hungary	NL	Netherlands
BE	Belgium	IT	Italy	NO	Norway
BG	Bulgaria	JP	Japan	RO	Romania
BR	Brazil	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic	KR	Republic of Korea	SE	Sweden
CG	Congo	LI	Liechtenstein	SN	Senegal
CH	Switzerland	LK	Sri Lanka	SU	Soviet Union
CM	Cameroon	LU	Luxembourg	TD	Chad
DE	Germany, Federal Republic of	MC	Monaco	TG	Togo
DK	Denmark	MG	Madagascar	US	United States of America
FI	Finland	ML	Mali		
FR	France				

VECTORS CONTAINING A GENE FOR A PROTEIN HAVING ERYTHROID-POTENTIATING ACTIVITY AND RECOMBINANT DNA METHODS FOR PRODUCING SAID PROTEIN

Field of the Invention

This invention relates to the production of a protein having erythroid-potentiating activity (EPA) and particularly to a pure EPA protein, to vectors containing the gene for expressing said protein, to microorganisms and mammalian cells transformed with said vectors, and to methods for producing said protein by recombinant DNA techniques.

Background of the Invention

The many different cell types found in blood are all derived from pluripotent hematopoietic stem cells. Stem cells perform two functions: (1) they reproduce themselves, thereby maintaining a stem cell population in the body and (2) they provide progeny committed to differentiate into any of the mature blood cell types. A cell which is committed to differentiate along a particular hematopoietic pathway is termed a progenitor cell. Progenitor cells for T lymphocytes, B lymphocytes, granulocytes, macrophages, red blood cells, platelets, and eosinophils, as well as earlier progenitors which can individually give rise to several of the mature cell types, have been studied experimentally both *in vivo* and *in vitro* (Dexter,

T.M. 1983 J. Pathology 141 415-433). It has been determined in vitro that proliferation and/or differentiation of each progenitor cell type depends upon "factors" which have been derived from various sources. For example, T lymphocytes require T-cell growth factor (IL-2) and some of the later progenitors of red blood cells require a factor called erythropoietin.

Several general classed of erythrocyte (red blood cell) progenitors have been distinguished by the time course of colony formation in vitro and the characteristics of the colonies produced. Although there is a continuum with respect to the maturity of erythroid colony-forming cells, the most mature of these is generally referred to as colony-forming unit - erythroid (CFU-E), giving rise to relatively small colonies consisting of 8 to 64 hemoglobinized cells. The burst-forming unit (BFU-E) is a more primitive erythroid progenitor. The BFU-Es require more time to mature than CFU-Es and they can give rise to very large colonies containing up to thousands of cells with multiple subcolonies. The CFU-E is thought to arise from the BFU-E. The CFU-E is sensitive to erythropoietin (EPO) whereas the BFU-E is relatively insensitive to EPO.

Several in vivo studies have suggested that erythropoietin does not affect the pluripotent hematopoietic stem cell or the early erythroid progenitor cell (BFU-E). As a result of these studies, Iscove postulated the existance of an independent modulator of multipotent stem cells and early committed erythroid cells. This hormone or group of hormones has been referred to as

burst-promoting activity (BPA) or erythroid-potentiating activity (EPA) because it is primarily assayed by its ability to stimulate the growth of BFU-Es. Studies of BPA have generally employed unpurified material, derived from various sources. One murine protein IL-3, which affects all myeloid progenitors including cells of the erythroid lineage as well as stem cells, has recently been purified to homogeneity and the cDNA for this protein has been cloned.

Common sources for EPA are peripheral blood mononuclear cells, bone-marrow conditioned medium, urine from anemic patients, serum, and normal and neoplastic cells of T-lymphocyte and mononuclear phagocyte lineage. The most widely used source of EPA is lectin-stimulated lymphocytes. Medium conditioned by phytohemagglutinin (PHA-), concanavalin-A- and pokeweed mitogen-stimulated lymphoid populations consistently produces EPA. Erythroid-potentiating activity is also released by lymphoid cells responding to antigen. It is most likely that in these circumstances the EPA is being elaborated by activated T lymphocytes. Consistent with this in vitro evidence for production of EPA by T lymphocytes are studies in which erythroid colony formation in vivo is greatly stimulated by addition of T lymphocytes.

Conditioned medium from heterogeneous cell populations has proved useful for biologic studies of the action of EPA; however, these materials have not been highly suitable for biochemical investigations. For this reason, the search for cell lines

producing EPA has been actively pursued. Several cell lines are known to produce EPA. In general, human T-cell lines composed of mature helper T cells, which elaborate other lymphokines, produce EPA. Undifferentiated T-cell lines, which are widely used in studies of T-cell function, do not appear to produce EPA. Two human monocyte-like cell lines, U-937 and GCT elaborate EPA.

The characterization and partial purification of a human T-lymphocyte-derived factor that stimulates erythroid progenitors in vitro was reported by Lusis and Golde in Proceedings of the Second Conference on Hemoglobin Switching, G. Stammatoyannopoulos and A.W. Nienhuis, eds., Gume & Stratton (1981). Their source of this factor (EPA) is a human T-lymphoblast cell line (Mo) that was derived from a patient with a T-cell variant of hairy cell leukemia. Medium conditioned by Mo cells in the absence of serum is a rich source of EPA, and using this source EPA was purified about 250-fold. The final product stimulated human CFU-E and BFU-E progenitors in vitor in nanogram quantities. Physically, EPA appears to be an acidic glycoprotein of molecular weight about 45,000, although it exhibits a degree of size and charge heterogeneity. EPA was separable from a granulocyte-macrophage colony-stimulating factor (G,M-CSF) also produced by Mo cells. However, the "purified" EPA retained an activity that stimulates the formation of mixed myeloid colonies in vitro, and may be employed to stimulate the formation and growth of myeloic cells including granulocytes, macrophages, megacaryocytes and mast cells.

Because of its stimulating activity on erythroid progenitors, EPA can be useful for the treatment of anemia associated with chronic renal failure and for adjunctive therapy

in patients receiving cancer or other chemotherapy. An adequate source of EPA will also be of great value in medical research to study the growth and formation of erythroid cells. EPA treatment can also be used for chronic diseases having associated anemia such as rheumatoid arthritis, tuberculosis, osteomyelitis, chronic urinary tract infections, and some malignancies. As in the anemia or chronic renal failure, EPA will be useful in treating the low red blood cell counts of patients with the above chronic diseases, however it will not treat the underlying basic disease process.

Because the effect of EPA will be to reduce the numbers of transfusions required by patients having chronic renal failure, the likelihood of contracting transfusion related infections which include hepatitis A, hepatitis B, non A and non B hepatitis, infectious mononucleosis, CMV disease, toxoplasmosis, malaria, syphillis and trypanosomiasis will be reduced.

Thus, it can be appreciated that it would be desirable to have a source for EPA that could readily supply this protein in quantities sufficient for the above uses. Recently developed techniques have made it possible to employ microorganisms or eukaryotic cells, capable of rapid and abundant growth, for the synthesis of commercially useful proteins and peptides, regardless of their source in nature. The general approach is to isolate (clone) the DNA sequence (gene) encoding the protein of interest, to insert that gene into a nucleic acid vector containing signals for abundant expression of the protein, and to

transfer the vector with inserted gene into a suitable host cell. This host-vector combination can be grown in volume to produce quantities of the protein, and the protein can then be recovered by known separation and purification techniques.

One method of gene isolation is based upon knowing part or all of the amino acid sequence of the protein encoded by the gene. Each of the twenty commonly occurring amino acids is encoded by one or more triplets of nucleotides called codons (for discussion see J.D. Watson, Molecular Biology of the Gene, (W.A. Benjamin, Inc., 1977, p. 347-377). Knowing the deoxyribonucleotide sequence of the gene (or DNA sequence which codes for a particular protein) allows the exact description of that protein's amino acid sequence. However, the converse is not true; for example, although tryptophan or methionine are coded for by only one codon, the other amino acids can be coded for by up to six codons. Thus, there is considerable ambiguity in predicting the nucleotide sequence from the amino acid sequence. However, a pool of synthetic oligonucleotides, corresponding to all possible combinations of codons for a particular sequence of amino acids within a protein, should include the nucleotide sequence which in fact occurs in the gene encoding that protein. Such synthetic oligonucleotide pools can be used to screen cloned gene libraries by standard techniques in order to identify the clone of interest.

Determination of part or all of the amino acid sequence of a protein, as a basis for synthesis of oligonucleotide screening

probes, requires that the protein be purified to homogeneity. In the case of EPA, prior to the present invention, only a partial purification had been achieved which was not sufficient to obtain the required amino acid sequence to synthesize suitable screening probes.

Summary of the Invention

The present invention overcomes these problems and provides a ready souce of EPA using recombinant DNA technology. Thus, the present invention provides a substantially pure EPA protein having a biological activity of at least about 10^6 units per mg of protein and having an apparent molecular weight of about 28,000 daltons as determined by SDS-PAGE analysis on 10 percent polyacrylamide gel, a cloned EPA gene coding for such EPA protein, a recombinant vector containing the EPA gene, a microorganism or mammalian cell transformed with such a recombinant vector, and a method for producing EPA by expressing said EPA gene by culturing said microorganism or mammalian cell.

Purified EPA stimulates the growth of both early and late erythroid precursors from human bone marrow, as well as colony formation by the K562 erythroleukemia cell line.

Brief Description of the Drawings

Fig. 1 is a schematic illustrating the preparation of plasmid pTPL from plasmid pAdD26SVpA(3).

Fig. 2 is a schematic continuing from Fig. 1 and

illustrating the preparation of plasmid p91023 from plasmid pTPL.

Fig. 3 is a schematic continuing from Fig. 2 and illustrating plasmid p91023(B).

Fig. 4 sets forth a nucleotide sequence containing an EPA gene cloned in accord with one embodiment of this invention along with the deduced amino acid sequence of the EPA protein that the gene encodes.

Fig. 5 sets forth another nucleotide sequence containing a sequence coding for EPA in accord with another embodiment of the invention along with the deduced amino acid sequence of the EPA protein encoded thereby.

Detailed Description of the Invention

The following definitions are supplied in order to facilitate the understanding of this case. To the extent that the definitions vary from meanings circulating within the art, the definitions below are to control.

Amplification means the process by which cells produce gene repeats within their chromosomal DNA.

Downstream means the direction going towards the 3' end of a nucleotide sequence.

An enhancer is a nucleotide sequence that can potentiate the transcription of genes independent of the identity of the gene, the position of the sequence in relation to the gene, or the orientation of the sequence.

A gene is a deoxyribonucleotide sequence coding for a given mature protein. For the purposes herein, a gene shall not include untranslated flanking regions such as RNA transcription initiation signals, polyadenylation addition sites, promoters or enhancers.

A selection gene is a gene that confers a phenotype on cells which express the gene as a detectable protein.

A selection agent is a condition or substance that enables one to detect the expression of a selection gene.

Phenotype means the observable properties of a cell as expressed by the cellular genotype.

Genotype means the genetic information contained within a cell as opposed to its expression, which is observed as the phenotype.

Ligation is the process of forming a phosphodiester bond between the 5' and 3' ends of two DNA strands. This may be accomplished by several well known enzymatic techniques, including blunt end ligation by T4 ligase.

Orientation refers to the order of nucleotides in a DNA sequence. An inverted orientation of a DNA sequence is one in which the 5' to 3' order of the sequence in relation to another sequence is reversed when compared to a point of reference in the DNA from which the sequence was obtained. Such points of reference can include the direction of transcription of other specified DNA sequences in the source DNA or the origin of replication of replicable vectors containing the sequence.

Transcription means the synthesis of RNA from a DNA template.

Transformation means changing a cell's genotype by the stable transfection of cells with exogenous DNA. Transformation may be detected in some cases by an alteration in cell phenotype.

Transformed cells are called transformants. Pre-transformation cells are referred to as parental cells.

Translation means the synthesis of a polypeptide from messenger RNA.

Erythroid-potentiating activity (EPA) can be derived from a number of cellular sources including peripheral blood mononuclear cells, bone-marrow conditioned medium, urine from anemic patients, serum, and normal and neoplastic cells of T-lymphocyte and mononuclear phagocyte lineage. Because EPA has not been purified to homogeneity and because the assay for EPA is complex, the amino acid sequence for EPA is unknown. Thus, an oligonucleotide probe for EPA mRNA could not be predicted with any degree of certainty from any literature sources.

However, there have been reports of partial purification of EPA. For instance, Lusis and Golde, supra, reported a 250-fold purification of EPA from the Mo cell line with the resulting product having a specific activity of about 51,000 units/mg protein and Porter and Ogawa, Blood, 59 p. 1207 (1982) have reported a 300-fold purification of EPA from human bone-marrow conditioned medium. Also Iscove at pp. 37-52 in Golde, et al., Eds., Hematopoietic Cell Differentiation. ICN-UCLA Symposia on Molecular and Cellular Biology, Vol. 10, Academic Press, New York (1978) reported the purification of EPA from mouse spleen cells, Dubes, et al, at pp. 119-128 in Golde, et al., Eds., supra, and in Exp. Hematol. 8 (Supp. 8) p. 128 (1980) reported the separation of EPA from human urine, and recently Abboud, et al.,

Blood, 58 pp. 1148 (1981) reported the use of hydrophobic adsorption chromatography to separate EPA from the colony-stimulating activity factor (CSA). In general, the activities that have been characterized appear to be glycoproteins of molecular weight 30,000 to 45,000 and the reported heat stabilities have varied considerably.

EPA was partially purified from Mo cell line conditioned medium in accord with the procedures described in Lusis and Golde, Proceedings of the Second Conference on Hemoglobin Switching, supra. The partially purified EPA would not permit an analysis sufficient to construct possible oligonucleotide probes for EPA. Thus, further purification to substantial homogeneity was required. Such purification was accomplished in accord with the present invention by sequential concentration, ammonium sulfate precipitation, lentil-lectin affinity chromatography, gel filtration and reverse phase high performance liquid chromatography (HPLC). EPA was assayed by its ability to stimulate the growth of large erythroid colonies (bursts) from normal human peripheral blood. The purified EPA in accord with the present invention has an apparent molecular weight of 28,000 and appears as a single band when analyzed by SDS-PAGE under reducing or nonreducing conditions.

The EPA protein was analyzed on a gas phase sequenator to obtain a paartial amino acid sequence which was sufficient to predict possible DNA sequences and make oligonucleotide probes. All sixty-four possible 12-mers that could encode residues 9-13

of the analyzed amino acid sequence, all sixty-four possible 14-mers that could encode residues 21-25 of the analyzed amino acid sequence, and two pools each of sixteen possible 53-mers (based on coding usage preferences) that could encode residues 9-26 of the analyzed amino acid sequence were prepared.

T-lymphocytes were induced with phytohemagglutinin (PHA) and phorbol myristic acid (PMA) to enhance their lymphokine production. The cells were centrifuged and cytoplasmic RNA was prepared by a standard gentle lysis procedure. Polyadenylated messenger RNA was then prepared by chromatography on oligo dT cellulose.

First strand cDNA was prepared using standard methods using the mRNA prepared above. To enrich for sequences specific for lymphokines, the first strand cDNA was hybridized with an excess of mRNA from cells that do not make lymphokines. Most of the sequences expressed in the T-lymphocytes that produce lymphokines will also be expressed in, for instance, B-cell lines and immature T-cell lines. A small fraction of sequences will be found only in the starting T-lymphocyte cell line. Most of the single stranded cDNA will anneal to the mRNA from the B-cell lines or immature T-cell lines and become double stranded. The cDNA unique to the starting T-lymphocyte cell line remains single stranded. The single and double stranded DNA are separated by chromatography on hydroxylapatite, thereby substantially concentrating the lymphokine sequences.

The single stranded cDNA was then converted to

double-stranded form by treatment with the Klenow fragment of DNA polymerase I using standard techniques. The loopback form was cleaved with S1-nuclease and C-tails were added using terminal transferase by standard methods. Finally, the C-tailed cDNA was annealed with G-tailed pBR322 and used to transform E. coli. The E. coli clones were spread onto agar plates and the colonies were lifted off with nitrocellulose filter disks.

After growing these colonies, replicas were made by carefully placing a pre-wetted filter on top of the original filter. The replica filters were aligned with the original by punching 3 asymmetrical holes through the filters with a needle. The replica filters were removed to fresh plates and grown. The plasmid DNA in each colony was amplified in situ by transferring the filters to fresh plates containing 100 ug/ml chloramphenicol and incubating overnight. Finally, these replica filters were prepared for standard colony hybridization by treatment with base, followed by neutralization and drying. The dried filters were baked to fix the liberated DNA to the nitrocellulose.

To identify EPA clones, one set of replica filters was probed with the pools of 12-mers and another corresponding set of duplicate filters was probed with the pools of 14-mers.

Potential EPA clones were identified by aligning the duplicate filters and looking for individual colonies which hybridized to both independent probes (the 64 14-mers and 64 12-mers). The clones which hybridized to both probes were picked by aligning the x-ray film of the hybridized filter with the

original master filter. These colonies were picked and grown in liquid culture for the preparation of analytical amounts of DNA (prepared by the standard base treatment method).

Samples of DNA from each of the picked clones were cleaved with PstI to excise the inserts. These digests were analyzed by agarose gel electrophoresis. A variety of insert sizes was observed. The gel was next prepared for standard Southern analysis. Nitrocellulose strips were placed on either side of the gel. This sandwich was placed on top of a 1 inch stack of paper towels and a second stack of towels was placed on top of the gel. DNA was allowed to leach out of the gel overnight (in both directions) giving replica filters of the gel. These filters were baked and then probed overnight as described for the colony hybridization (1 filter with the 12-mers and 1 with 14-mers). After washing the filters and exposing them to film it was observed that three clones (designated 27, 57, 58) hybridized most strongly to both probes. These three clones all clearly had at least one internal PstI site suggesting that they might encode the same sequence. To further test the clones, the Southern filters were washed to eliminate the radioactivity and were reprobed with the 2 pools of labeled 53-mers. Both sets of 53-mers hybridized strongly only with the three clones designated 27, 57, and 58 suggesting that these clones encoded the EPA sequence above.

Each of the 8 pools of 14-mers was tested separately to determine which pool hybridized most strongly with the EPA

clones. Pool 14-5 proved to hybridize most strongly to clones 27, 57, and 58 indicating that this pool of oligonucleotides would be a good primer for dideoxy sequencing of the candidate clones.

DNA from each candidate clone was purified using cesium chloride equilibrium density centrifugation and sequenced using standard dideoxy sequencing with pool 14-5 as the primer. The complete nucleotide sequence was obtained by subcloning the PstI fragments of the clone (the EPA gene was found to have 3 internal PstI sites which generate four small fragments for sequencing) into M13 and sequencing them using the universal M13 primer. To align the sequences of the four PstI fragments, oligonucleotide primers were synthesized for use in sequencing the plasmid DNA directly (as described above with pool 14-5). One of these primers

d(T-G-C-A-C-C-T-G-T-G-T-C-C-C-A-C-C-)

encoded the end of the mature protein and could be used for insertion into the bacterial expression vector as well as sequencing at the 5' end of the coding region. A second primer;

d(G-A-G-G-A-G-T-T-T-C-T-C-A-T-T-G-C-T-G)

encoded the region near the 3' end of the first PstI fragment and could be used to sequence through the three PstI sites towards the 3' end of the clone. The final primer;

d(C-C-A-C-A-A-G-C-A-A-T-G-A-G-G-T-G-C-C-)

was complementary to the mRNA strand near the 5' end of the 3' Pst fragment of the clone and could be used to sequence toward

the 5' end of the clone through the 3 PstI sites. The complete DNA sequence of the EPA gene of clone 57 is illustrated in Fig. 4 along with the amino acid sequence of the translated protein product.

Once the full length EPA cDNA clone is obtained, known and appropriate means are utilized to express the EPA protein, e.g. insertion into an appropriate vector, transfection into an appropriate host, selection of transformants, and culturing these transformants to express EPA activity.

Host-vector systems for the expression of EPA may be procaryotic, but the complexity of EPA may make the preferred expression system a mammalian one. This is easily accomplished by transformation or by microinjection of eucaryotic (usually mammalian or vertebrate) cells with a suitable EPA vector. Eucaryotic transformation is in general a well-known process, and may be accomplished by a variety of standard methods. These include the use of protoplast fusion, chromosome transfection, lytic and nonlytic viral vectors (For example, Mulligan et al., "Nature" (London) 277:108-114 [1979], cell-cell fusion (Fournier et al., "Proc. Nat. Acad. Sci." 74:319-323 [1977], lipid structures (U.S. patent 4,394,448) and cellular endocytosis of DNA precipitates (Bachetti et al., "Proc. Nat. Acad. Sci. 74:1590-1594 [1977]).

Transformation which is mediated by lytic viral vectors is efficient but is disadvantageous for a number of reasons: the maximum size of transfected DNA is limited by the geometry of

viral capsid packing, the exogenous genes are frequently deleted during viral replication, there is a requirement for helper virus or specialized hosts, host cells must be permissive, and the hosts are killed in the course of the viral infection.

Nonlytic transformations are based on the transcription and translation of virus vectors which have been incorporated into a cell line as a stable episome. These systems generally require unique cell lines and suffer from a number of disadvantages. See "Trends in Biochemical Sciences", June 1983, pp. 209-212.

On the other hand, other transformation methods in which extrachromosomal DNA is taken up into the chromosomes of host cells have been characterized by low frequencies of transformation and poor expression levels. These initial difficulties were ameliorated by transformation with genes which inheritably confer selectable phenotypes on the small subpopulation of cells that are in fact transformed (selection genes). The entire population of transformed cells can be grown under conditions favoring cells having acquired the phenotype, thus making it possible to locate transformants conveniently. Thereafter, transformants can be screened for the capability to more intensely express the phenotype. This is accomplished by changing a selection agent in such a way as to detect higher expression.

Selection genes fall into three categories: Detectably amplified selection genes, dominant selection genes, and detectably amplified dominant selection genes.

Detectably amplified selection genes are those in which amplification can be detected by exposing host cells to changes in the selection agent. Detectably amplified genes which are not dominant acting generally require a parental cell line which is genotypically deficient in the selection gene. Examples include the genes for hydroxymetholglutanyl CoA reductase (Sinensky, "Biochem. Biophys. Res. Commun" 78:863 [1977], ribonucleotide reductase (Meuth et al. "Cell:3:367 [1943], aspartate transcarbamylase; (Kemp et al. "Cell" 9:541 [1976], adenylate deaminase (DeBatisse et al. "Mol and Cell Biol." 2(11):1346-1353 [1982] mouse dihydrofolate reductase (DHFR) and, with a defective promoter, mouse thymidine kinase (TK).

Dominant selection genes are those which are expressed in transformants regardless of the genotype of the parental cell. Most dominant selection genes are not detectably amplified because the phenotype is so highly effective in dealing with the selection agent that it is difficult to discriminate among cell lines that have or have not amplified the gene. Examples of dominant selection genes of this type include the genes for procaryotic enzymes such as xanthine-guanine phosphoribosyltransferase (Mulligan et al. "Proc. Nat. Acad. Sci." 78[4]:2072-2076 [1981] and aminoglycoside 3' - phosphotransferase (Colbere-Garapin et al., "J. Mol. Biol.", 150:1-14 [1981].

Some dominant selection genes also are detectably amplified. Suitable examples include the mutant DHFR gene described by Haber

et al., "Somatic Cell Genet." 4:499-508 (1982), cell surface markers such as HLA antigens and genes coding for enzymes such as specific esterases that produce fluorescent or colored products from fluorogenic or chromogenic substrates as is known in the art.

Detectably-amplified, dominant selection genes are preferred for use herein. It should be understood that a dominant selection gene in some cases can be converted to a detectably amplified gene by suitable mutations in the gene.

Selection genes at first were of limited commercial utility. While they enabled one to select transformants having the propensity to amplify uptaken DNA, most selection genes produced products of no commercial value. On the other hand, genes for products which were commercially valuable generally did not confer readily selectable (or even detectable) phenotypes on their transformants. This would be the case, for example, with enzymes or hormones which do not provide transformed cells with unique nutrient metabolic or detoxification capabilities. Most proteins of commercial interest fall into this group, e.g. hormones, proteins participating in blood coagulation and fibrinolytic enzymes.

Thus, vectors used in EPA transformation can contain a selection gene and the EPA gene. In addition, there usually will be present in the transformation vectors other elements such as enhancers, promoters, introns, accessory DNA, polyadenylation sites and 3' noncoding regions as will be described below.

Suitable selection genes are described above. It is preferred that the selection agent be one that prevents cell growth in the absence of the selection gene. That way, revertant cells in large scale culture that lose the selection gene (and presumably the EPA gene as well) will not over-grow the fermentation. However, it would be desirable in the commercial production of EPA to avoid the use of cell toxins, thereby simplifying the product purification steps. Thus, a desirable selection gene would be one that enables transformants to use a nutrient critical for growth that they otherwise would not be able to use. The TK gene described above is an example.

Two classes of vectors can be employed in transformation. The first class are the unlinked vectors. Here the selection gene and the EPA gene are not covalently bound. This vector class is sometimes preferred because the step of ligating or otherwise bonding the two genes is not required. This simplifies the transformation process because the selection and product genes usually are obtained from separate sources and are not ligated in their wild-type environment. In addition, the molar ratio of the EPA and selection genes employed during transformation can be adjusted to increase transformation efficiency.

The second class of vectors is linked vectors. These vectors are distinguished from unlinked vectors in that the selection and EPA genes are covalently bound, preferably by ligation.

The vectors herein may also include enhancers. Enhancers are functionally distinct from promoters, but appear to operate in concert with promoters. Their function on the cellular level is not well understood, but their unique characteristic is the ability to activate or potentiate transcription without being position or orientation dependent. Promoters need to be upstream of the gene, while enhancers may be present upstream or 5' from the promoter, within the gene as an intron, or downstream from the gene between the gene and a polyadenylation site or 3' from the polyadenylation site. Inverted promoters are not functional, but inverted enhancers are. Enhancers are cis-acting, i.e., they have an effect on promoters only if they are present on the same DNA strand. For a general discussion of enhancers see Khoury et al., "Cell" 33:313-314 (1983).

Preferred enhancers are obtained from animal viruses such as simian virus 40, polyoma virus, bovine papilloma virus, retrovirus or adenovirus. Ideally, the enhancer should be from a virus for which the host cell is permissive, i.e. which normally infects cells of the host type. Viral enhancers may be obtained readily from publically available viruses. The enhancer regions for several viruses, e.g., Rous sarcoma virus and simian virus 40, are well known. See Luciw et al., "Cell" 33:705-716 (1983). It would be a matter of routine molecular biology to excise these regions on the basis of published restriction maps for the virus in question and, if necessary, modify the sites to enable splicing the enhancer into the vector as desired. For example,

see Kaufman et al, "J. Mol. Biol.", 159:601-621 (1982) and "Mol. Cell Biol." 2(11):1304-1319 (1982). Alternatively, the enhancer may be synthesized from sequence data; the sizes of viral enhancers (generally less than about 150 bp) are sufficiently small that this could be accomplished practically.

Another element which should be present in the vector assembly is a polyadenylation splicing (or addition) site. This is a DNA sequence located downstream from the translated regions of a gene, shortly downstream from which in turn transcription stops and adenine ribonucleotides are added to form a polyadenine nucleotide tail at the 3' end of the messenger RNA. Polyadenylation is important in stabilizing the messenger RNA against degradation in the cell, an event that reduces the level of messenger RNA and hence the level of product protein.

Eucaryotic polyadenylation sites are well known. A concensus sequence exists among eucaryotic genes: The hexanucleotide 5'-AAUAAA-3' is found 11-30 nucleotides from the point at which polyadenylation starts. DNA sequences containing polyadenylation sites may be obtained from viruses in accord with published reports. Exemplary polyadenylation sequences can be obtained from mouse beta-globin, and simian virus 40 late or early region genes, but viral polyadenylation sites are preferred. Since these sequences are known, they may be synthesized in vitro and ligated to the vectors in conventional fashion.

A polyadenylation region must be located downstream from the

EPA gene. The sequence which separates the polyadenylation site from the translational stop codon is preferably an untranslated DNA oligonucleotide such as an unpromoted eucaryotic gene. Since such oligonucleotides and genes are not endowed with a promoter they will not be expressed. The oligonucleotide should extend for a considerable distance, on the order of up to about 1,000 bases, from the stop codon to the polyadenylation site. This 3' untranslated oligonucleotide generally results in an increase in product yields. The vector may terminate from about 10 to about 30 bp downstream from the concensus sequence for polyadenylation, but it is preferable to retain the 3' sequences found downstream from this site in its wild-type environment. These sequences typically extend about from 200 to 600 base pairs downstream from the polyadenylation site.

The vectors described herein may be synthesized by techniques well known to those skilled in this art. The components of the vectors such selection genes, enhancers, promoters, and the like may be obtained from natural sources or synthesized as described above. Basically, if the components are found in DNA available in large quantity, e.g. components such as viral functions, or if they may be synthesized, e.g. polyadenylation sites, then with appropriate use of restriction enzymes large quantities of vector may be obtained by simply culturing the source organism, digesting its DNA with an appropriate endonuclease, separating the DNA fragments, identifying the DNA containing the element of interest and

recovering same. Ordinarily, a transformation vector will be assembled in small quantity and then ligated to a suitable autonomously replicating synthesis vector such as a procaryotic plasmid or phage. The pBR322 plasmid may be used in most cases. See Kaufman et al., op. cit.

The synthesis vectors are used to clone the ligated transformation vectors in conventional fashion, e.g. by transfection of a permissive procaryotic organism, replication of the synthesis vector to high copy number and recovery of the synthesis vector by cell lysis and separation of the synthesis vector from cell debris.

The resulting harvest of synthesis vector may be directly transfected into eucaryotic cells, or the transformation vector may be rescued from the synthesis vector by appropriate endonuclease digestion, separation by molecular weight and recovery of the transformation vector. Transformation vector rescue is not necessary so long as the remainder of the synthesis vector does not adversely affect eucaryotic gene amplification, transcription or translation. For example, the preferred synthesis vector herein is a mutant of the E. coli plasmid pBR322 in which sequences have been deleted that are deleterious to eucaryotic cells. See Kaufman et al., op. cit. Use of this mutant obviates any need to delete the plasmid residue prior to transformation.

The cells to be transformed can be any prokaryotic cell or any eucaryotic cell, including yeast protoplasts and insect

cells, but ordinarily will be a nonfungal cell. Candidate cells need not be genotypically deficient in the selection gene so long as the selection gene is dominant acting.

Eucaryotic cells preferably will be stable mammalian cell lines as is discussed above. Cell lines that are known to stably integrate selection genes into their chromosomal DNA are best, for example Chinese hamster ovary (CHO) cell lines. Also useable are HeLa, COS monkey cells, melanoma cell lines such as the Bowes cell line, mouse L cells, mouse fibroblasts and mouse NIH 3T3 cells.

Methods for facilitating cellular uptake of DNA are described above. Microinjection of the vector into the cell nucleus will yield the highest DNA uptake efficiencies, but exposing parental cells to DNA in the form of a calcium phosphate precipitate is generally most convenient.

The population of cells that has been exposed to transforming conditions is then processed to identify the transformants. Only a small subpopulation of any culture which has been treated for transformation will exhibit the phenotype of the selection gene. The cells in the culture are screened for the phenotype. This can be accomplished by assaying the cells individually with a cell sorting device where the phenotype is one that will produce a signal, e.g. fluorescence upon cleavage of a fluorogenic substrate by an enzyme produced by the selection gene. Preferably, however, the phenotype enables only transformants to grow or survive in specialized growth media as

is further discussed above.

Selection transformants then will be screened for ligation of the product gene into their chromosomes or for expression of the product itself. The former can be accomplished using Southern blot analysis, the latter by standard immunological or enzymatic assays.

Once the tranformants have been identified, steps are taken to amplify expression of the product gene by further cloning in the presence of a selection agent such as methotrexate (MTX).

The DNA sequence coding for a protein exhibiting EPA activity in accord with the present invention, such as illustrated in Fig. 4, can be modified by conventional techniques to produce variations in the final EPA protein which still have EPA activity in the assay tests described herein. Thus, for example, one, two, three, four or five amino acids can be replaced by other amino acids. Belgian Patent No. 898,016, which is incorporated herein by reference, describes one such typical technique for replacing cysteine by e.g., serine.

EPA/cDNA in accord with this invention includes the mature EPA/cDNA gene preceded by an ATG codon and EPA/cDNA coding for allelic variations of EPA protein. One allele is illustrated in Fig. 4. The EPA protein of this invention includes the l-methionine derivative of EPA protein (Met-EPA) and allelic variations of EPA protein. The mature EPA protein illustrated by the sequence in Fig. 4 begins with the sequence Ala.Pro.Phe.Glu... the beginning of which is depicted by a arrow

after nucleotide number 44 in Fig. 4. The Met-EPA would begin with the sequence Met.Ala.Pro.Phe.Glu... EPA protein of the present invention exhibits a specific activity of at least 10^6 units/mg of protein.

As aforesaid, the EPA protein in accord with the present invention can be used for treatment of patients having low red blood cell counts to stimulate growth and formation of erythroid cells. For such use, a daily dosage of about 200 to 1000 ug per patient is typically indicated. The EPA protein is preferably injected into the patient intravenously in a suitable pharmacological carrier. Examples of such carriers include pharmacological saline and human serum albumin in saline.

Hence, embodiments of the invention and subject matter provided thereby include, without limitation: 1) a method for preparing an expression vector containing an EPA gene or DNA sequence coding for EPA, said method being as generally outlined and specifically described herein; 2) cells transformed with an expression vector prepared according to the method indicated in 1), above, including 2-A) prokaryotic cells, 2-B) eukaryotic cells and 2-C) mammalian cells; 3) cloned DNA sequences which code for EPA including the sequences 3-A) set forth in Fig. 4 and 3-B) set forth in Fig. 5); 4) Recombinant expression vectors comprising a DNA sequence which codes for EPA protein; 5) transformed cells comprising an expression vector as indicated in

4), above, including 5-A) prokaryotic cells, 5-B) eukaryotic cells, 5-C) yeast cells and 5-D) mammalian cells; 6) EPA protein made by expressing a recombinant expression vector containing an EPA gene or DNA sequence coding therefor in a transformed cell including 6-A) prokaryotic cells, 6-B) eukaryotic cells, 6-C) yeast cells and 6-D) mammalian cells; 7) Human EPA protein substantially free of other protein of human origin; 8) Gibbon ape EPA protein substantially free of other protein of Gibbon origin; 9) EPA protein substantially free of glycosylation; 10) EPA protein glycosylated by expression of a recombinant expression vector containing an EPA gene or DNA sequence coding therefor in a transformed eukaryotic cell; 11) EPA protein having an activity of at least about 10^6 units per milligram of protein including 11-A) protein of human origin, 11-B) protein of Gibbon ape origin, 11-C) EPA protein having the amino acid sequence set forth in Fig. 4 or an allelic variation thereof and 11-D) EPA protein having the amino acid sequence set forth in Fig. 5 or an allelic variation thereof; 12) cDNA which codes for EPA including 12-A) cDNA having the nucleotide sequence illustrated in Fig. 4 and 12-B) cDNA having the nucleotide sequence illustrated in Fig. 5; 13) expression vectors comprising cDNA coding for EPA including 13-A) those comprising the nucleotide sequence illustrated in Fig. 4 and 13-B) those comprising the nucleotide sequence illustrated

in Fig. 5; 14) cells transformed with an expression vector as indicated in 13, 13-A and 13-B, above, including in each case 14-A) prokaryotic cells, 14-B) eukaryotic cells, 14-C) yeast cells, 14-D) mammalian cells and 14-E) insect cells; 15) EPA protein made by expressing cDNA coding for EPA in a transformed cell including 15-A) prokaryotic cells, 15-B) eukaryotic cells, 15-C) yeast cells, 15-D) mammalian cells and 15-E) insect cells; 16) EPA protein substantially free of other protein of native origin; 17) EPA protein glycosylated by expression of cDNA coding for EPA in a transformed eukaryotic cell including 17-A) mammalian cell and 17-B) insect cells; 18) EPA protein made by expressing in a cell cDNA having the nucleotide sequence illustrated in Fig. 4 and Fig. 5 including expression in each case in 18-A) prokaryotic cells, 18-B) eukaryotic cells, 18-C) yeast cells, 18-D) mammalian cells and 18-E) insect cells; 19) purified EPA protein conforming to the amino acid sequence Met-EPA including 19-A) such protein having an activity of at least about 10^6 units per milligram of protein; 20) purified EPA protein which is a mixture of EPA and Met-EPA including 20-A) such protein having an activity of at least about 10^6 units per milligram of protein; 21) therapeutic compositions for the treatment of mammals, particularly primates, to stimulate the formation and growth of erythroid cells comprising erythroid cell formation and growth stimulating

effective amount of EPA protein in a pharmaceutically acceptable carrier or diluent; 22) methods for the treatment of mammals having low red blood cell counts comprising administering to said mammal an effective amount of EPA protein including 22-A) administering said effective amount intravenously such as by injection in the form of a therapeutic composition comprising EPA protein in a pharmaceutically acceptable carrier; 23) methods for treating anemia associated with chronic renal failure comprising administering an effective amount of EPA protein; and 24) method for treating as adjunctive therapy the adverse effects including reduced red blood cell counts or anemia of chemotherapy including 24-A) cancer chemotherapy, in patients receiving chemotherapy causing such adverse effects.

The invention will be further understood with reference to the following illustrative embodiments, which are purely exemplary, and should not be taken as limitive of the true scope of the present invention, as described in the claims.

Unless otherwise noted, restriction endonucleases are utilized under the conditions and in the manner recommended by their commercial suppliers. Ligation reactions are carried on as described by Maniatis, T. et al., Molecular Cloning - A Laboratory Manual Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 1982 p. 245-6, the disclosure of which is incorporated herein by reference, using the buffer described at page 246

thereof and using a DNA concentration of 1-100 ug/ml, at a temperature of 23°C for blunt ended DNA and 16°C for "sticky ended" DNA. "Phosphatasing" as described herein, refers to dephosphorylation of DNA, and is carried out in the manner described by Maniatis et al., supra, e.g. at page 133 et seq. "Kinasing" refers to phosphorylation of DNA. Electrophoresis is done in 0.5-1.5% Agarose gels containing 90 mM Tris-borate, 10 mM EDTA. "Nick-translating" refers to the method for labeling DNA with 32p as described by Rigby et al., J. Mol. Biol., 113:237 (1977). All radiolabeled DNA is labeled with 32p, whatever labelling technique was used.

By "rapid prep" is meant a rapid, small scale production of bacteriophage or plasmic DNA, e.g., as described by Maniatis et al., supra, at p. 365-373.

In the examples, unless otherwise specified, temperatures are in °C.

EXAMPLE 1

Step A. Culture Of The Mo Cells

To obtain large volumes of Mo cells, cells were grown in T flasks containing about 200 ml of culture medium. Lots of 10 to 20 liters were prepared. A variant of the original Mo T-cell line was used to prepare the conditioned medium because these cells are able to grow in the absence of bovine serum (Chen, I.S., et al 1983 Nature 305: 502-505). Cells were grown for expansion at 37°C in Iscove's modified Dulbecco's medium (Irvine Scientific) supplemented with 0.1 percent bovine serum albumin (Irvine Scientific, selected lots), 0.035 percent dextrose, 2.2 uM human transferrin (Sigma) and antibiotics, to a final density

of about 10^6 per ml. The cells were centrifuged, and transferred to twice the original volume in a medium consisting only of Iscove's modified Dulbecco's medium containing antibiotics. They were incubated for an additional week, and then removed by centrifugation. After collection the medium was heated to 56°C for 30 min to inactivate the HTLV-II.

Step B. Biological Assays For EPA Activity

One method used to assay EPA is a modification of the previously described BFU assay. See Hunt et al, J. Biol. Chem. 256, pp. 7042-5 (1981). The peripheral blood leukocyte fraction (buffy coat) obtained from healthy donors was plated in a mixture consisting of 0.8 percent methylcellulose (Dow Chemical Co.,), 3.0 percent fetal bovine serum (Irvine Scientific and Hyclone selected lots), 0.5 U/ml of human urinary erythropoietin (supplied by the National Heart, Lung and Blood Institute about 1,140 U/mg protein), 10^{-6}M $FeCl_3$, 4.4 uM human transferrin, 1000 U/ml penicillin and 1000 ug/ml streptomycin, 10^{-4}M alpha-thioglycerol, 10^{-7}M selenium, 0.035 percent dextrose and Iscove's medium. The assays were done in 96-well tissue culture plates (Costar); each well contained 0.1 ml of methylcellulose mixture with cells at a density of 4 to 5 x 10^5 per ml, and 0.01 ml of test material. The material to be tested was dissolved in Dulbecco's phosphate-buffered saline (PBS; Irvine Scientific) containing 0.01 percent bovine serum albumin and sterile filtered. Two-fold serial dilutions were made; each dilution was

assayed in duplicate. Erythroid bursts were scored at 7 to 10 days using an inverted microscope. These bursts contain a minimum of 50 cells and are reddish brown in color indicating hemoglobinization of cells. Typically, control wells contained approximately 30 erythroid bursts per well. The last dilution which gave stimulation greater than 30 percent above control was considered to have 1 unit/ml of EPA.

BFU-E were also assayed from normal human bone marrow by a modification of a previously described technique, which employs microtiter dishes (Hunt, et al, supra); light density bone marrow cells were plated at 3 x 10^5 cells/ml in Iscove's medium containing 10 percent fetal bovine serum, 1 U/ml of human erythropoietin, 0.8 percent methylcellulose and antibiotics. Human bone marrow CFU-E were enumerated at 7 days from plates containing 2 x 10^5 bone marrow cells/plate (Hunt, R. et al 1981 J. Biol. Chem. 256: 7042-7045). Colony formation by K562 erythroleukemia cells was analyzed as described by Gauwerky, C. E., et al 1982 Blood 59: 300-305.

Step C. Purification Of EPA

All steps except HPLC were carried out at 4°C. After lentil lectin chromatography, manipulations were performed in polypropylene tubes or siliconized glassware to minimize loss by adsorption. Protein concentration was measured by the Bio-Rad method (Bio-Rad Laboratories).

Phenylmethylsulfonyl fluoride was added to 10 liters of

conditioned medium at a final concentration of 1 mM to inhibit proteolysis. The medium was then concentrated 15-fold with an Amicon hollow-fiber apparatus, using a Type H1P10-8 filter. The concentrate was brought to 80 percent saturation by the addition of solid ammonium sulfate. The precipitate was collected after stirring on ice for 4 to 12 hours, dialyzed against PBS, and clarified by centrifugation. This material was applied to a column (2.6 cm x 6 cm) of lentil lectin Sepharose 4B (Pharmacia) that had been equilibrated with PBS. The column was then washed with two column volumes of buffer, and EPA was eluted with 15 ml of 0.5 M alpha-methyl-D-mannoside in PBS. The eluate was collected and concentrated in an amicon stirred ultrafiltration cell, using a PM10 membrane, to a volume of 6 ml. The material was then applied to an Ultrogel ACA 44 column (1.6 x 80 cm; LKB) equilibrated with PBS as the running buffer. The flow rate was 6 ml per hour, and fractions of 2.5 ml were collected. Fractions with the highest specific activity were combined and applied directly to the HPLC column without additional manipulation. Reverse phase high performance liquid chromatography, HPLC, was performed on a Beckman apparatus, using a Vydac C-18 (10 x 250 mm) column with a starting solvent of 10 percent acetonitrile and 0.1 percent trifluoroacetic acid in water. The column was eluted with increasing concentrations of acetonitrile in trifluoroacetic acid using the following gradient: 10 percent to 30 percent acetonitrile over 10 min. then 30 percent to 60 percent over 60 min. The flow rate was 1.5 ml per min. Ultraviolet absorbance

was monitored at 220 nm, and protein peaks were collected manually, evaporated to dryness, and resuspended in a small volume of PBS.

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), was performed according to the method described by Laemmli, Nature 227, pp. 680-5 (1970). Proteins were visualized by iodination prior to electrophoresis. Iodinations were performed using a modification of the chloramine T method described by Hunter, et al, Nature 194, pp. 495-6 (1962); after reaction with ^{125}I and chloramine T the samples were not desalted by gel filtration but were applied directly to the gel. The tracking dye was electrophoresed to the end of the gel and excess ^{125}I migrated as the dye front. This procedure obviated the need for the addition of carrier protein and allowed complete recovery of small amounts of protein. The gels were fixed in 50 percent methanol, 10 percent acetic acid, soaked in 2 percent glycerol, dried, and exposed to film for 1 to 12 hours.

EPA from the earlier purification steps (ammonium sulfate precipitation and lentil lectin eluate) was inhibitory at moderate concentrations. Because of these considerations it was necessary to assay EPA over a wide concentration range for accurate quantitation. The addition of 0.01 percent bovine serum albumin to the dilutions was necessary to prevent loss by nonspecific adsorption, especially during sterile filtration. Bovine serum albumin has a slight stimulatory effect in the assay; however, this effect was negligible at the concentration

(0.01 percent) used.

It was difficult to accurately measure the amount of EPA in crude material and during the early steps of purification because of the presence of inhibitors of BFU-E growth. Although phenylmethylsulfonyl fluoride was routinely added to inhibit proteolysis, it may be unnecessary because the yields did not appear lower when it was omitted.

The overall purification resulted in an apparent yield of 25 percent; however, this figure is likely an overestimation due to inhibitors present in the crude material. Generally, approximately 25 ug is recovered from 10 liters of serum-free medium. The final product had a specific activity of 10^6 U/mg protein indicating that it is biologically active in the picomolar range. EPA retains activity over several months when stored at 4°C.

Highly purified ^{125}I-labelled EPA was subjected to SDS-PAGE analysis on 10 percent polyacrylamide gels. The protein migrates as a single broad protein band (characteristic of many glycoproteins) at a molecular weight of 28,000 in the presence or absence of 10 percent beta-mercaptoethanol, suggesting that it exists as a single polypeptide chain. A single band was observed with either silver staining or autoradiography of ^{125}I-labelled material. Attempts to elute the activity from SDS-PAGE gels by the method of Burgess et al, J. Biol. Chem. 252, pp. 1998-2003 (1977) yielded ambiguous results. Although activity appeared to coincide with the main protein fraction, the assays were

difficult to interpret because of the extreme variability of background activity due to the presence of BFU-E inhibitors in the gel eluates. When highly purified biologically active EPA was analyzed again by reverse phase HPLC, a single protein peak was observed. Finally, approximately 10 ug of this purified material was analyzed in an automated microsequenator and 11 cycles revealed single amino acids, confirming the presence of homogeneously purified protein.

Step D. Synthesis Of Probes

Two preparations of EPA containing about 25 ug of protein were analyzed on a gas phase sequenator. From analyses of four separate samples, we deduced a partial EPA amino acid sequence as follows:

1		5		10		15		20		25
(A) T	C	C	P	P	H	?	Q	T	A	F	C	N	S	D	L	V	I	R	A	K	F	V	G	T

From this sequence three different oligonucleotide probes were prepared.

The first probe consisted of all sixty-four possible 12-mers that encode residues 9-13 of the above sequence (i.e. Q T A F). These oligonucleotides were synthesized in 8 pools each containing 8 sequences that would be complementary to the EPA mRNA sequence as follows:

12-1 d(A-A-A-A-G-C-A-G-T-T-T-G)
G G G

12-2 d(A-A-A-T-G-C-A-G-T-T-T-G-)
G G G

12-3 d(A-A-A-A-G-C-T-G-T-T-T-G)
G G C

12-4 d(A-A-A-T-G-C-T-G-T-T-T-G)
G C C

12-5 d(A-A-A-A-G-C-A-G-T-C-T-G)
G G G

12-6 d(A-A-A-T-G-C-A-G-T-C-T-G)
G C G

12-7 d(A-A-A-A-G-C-T-G-T-C-T-G)
G G C

12-8 d(A-A-A-T-G-C-T-G-T-C-T-G)
G C C

The second probe consisted of all sixty-four 14-mers predicted from the sequence of residues 21-25 (AKFVG). Again, these compounds were synthesized in 8 pools of 8 sequences complementary to the mRNA sequence:

14-1 d(C-C-T-A-C-A-A-A-T-T-T-T-T-G-C)
C G C

14-2 d(C-C-A-A-C-A-A-A-T-T-T-T-G-C)
G G C

14-3 d(C-C-T-A-C-A-A-A-T-T-T-C-G-C)
C G C

14-4 d(C-C-A-A-C-A-A-A-T-T-T-C-G-C)
G G C

14-5 d(C-C-T-A-C-A-A-A-T-T-T-G-G-C)
C G C

14-6 d(C-C-A-A-C-A-A-A-T-T-T-G-G-C)
G G C

14-7 d(C-C-T-A-C-A-A-A-T-T-T-A-G-C)
C G C

14-8 d(C-C-A-A-C-A-A-A-T-T-T-A-G-C)
G G C

For the third probe, two pools of sixteen 53-mers were synthesized in which many of the possible sequences were eliminated by making guesses based on codon usage preferences (Mechanism of Aging and Development 118(4):285-314 (1982)). The 53-mers covered the region from residue 9 to residue 26. One pool (53-1) was synthesized to be complementary to the mRNA, while the other (53-2) was synthesized to have the same sense as the mRNA. The sequences were as follows:

53-1 d(G-T-G-C-C-C-A-C-A-A-C-T-T-A-G-C-C-C-G-G-A-T-G-A-C-C-A-G-
G T C

G-T-C-A-G-A-G-T-T-A-C-A-G-A-A-G-G-C-A-G-T-C-T-G)
G

53-2 d(C-A-G-A-C-T-G-C-C-T-T-C-T-G-T-A-A-C-A-G-T-G-A-T-C-T-G-G-
C C

T-G-A-T-C-C-G-G-G-C-T-A-A-G-T-T-T-G-T-G-G-G-C-A-C)
A C

Step E. Preparation Of cDNA Library

Mo cells were induced for 24 hr with PHA and PMA to enhance their lymphokine production (2 x 10^9 cells at 10^6 cells/ml in RPMI 1640 + 5% FCS). The cells were collected by centrifugation and cytoplasmic RNA was prepared by a standard gentle lysis procedure. Polyadenylated messenger RNA was prepared by chromatography on oligo dT cellulose.

First strand cDNA was prepared using standard methods beginning with 10ug of mRNA. The mRNA was removed by base

treatment and the cDNA was purified by gel exclusion chromatography (Sepharose CL-4B) and concentrated by ethanol precipitation. The yield was approximately 1 ug.

To try to enrich for sequences specific for Mo cells, the 1st strand cDNA was hybridized in 10ul with an excess of mRNA isolated from cells which do not make lymphokines. The cell lines used were Daudi (a B-cell line) and CEM (an immature T-cell line). Most of the sequences expressed in Mo, CEM and Daudi will be the same and only a small fraction will be found only in Mo. Therefore, by hybridizing 10 ug of CEM mRNA and 10 ug of Daudi mRNA to 1 ug of Mo cDNA we expected that most of the cDNA would anneal to its complementary mRNA and become double-stranded. The Mo lymphokine cDNAs should remain single-standed. Since double and single stranded molecules can be resolved by chromatography on hydroxylapatite, a substantial enrichment for lymphokine sequences should be achieved. After 18-24 hrs at 65°, the RNA:cDNA hybridization was passed over a 1 ml hydroxylapatite column. The flow through of the column was collected (the single-strand fraction) and used to make a cDNA library. Roughly 5% of the input cDNA was recovered. More than 90% was retained on the column (double-stranded) suggested that a 10-20 fold enrichment had been achieved.

The single-stranded cDNA was concentrated by butanol extraction, passed over Sepharose Cl-4B to remove phosphate buffer and collected by ethanol precipitation. The cDNA was then converted to a double-stranded form by treatment with the Klenow

fragment of DNA polymerase I using standard methods. The "loopback" form was cleaved with S1-nuclease and "C-tails" added with terminal transferase by standard methods. Finally, the tailed cDNA was annealed with G-tailed pBR322 (at the PstI site) and used to transform E. coli. From a final yield of 30ng of double stranded cDNA, approximately 40,000 clones were obtained which were spread onto 75 10 cm agar plates. The colonies were lifted off of the plates by placing pre-wetted nitrocellulose filter disks on top of them and removing the filters (colonies facing up) to fresh filters. After growing these colonies for 1 hour at 37°, 2 replicas were made by carefully placing a pre-wetted filter on top of the original filter. The replica filters were aligned with the original by punching 3 asymmetrical holes through the filters with a needle. The replica filters were removed to fresh plates and grown for several hours at 37°. The plasmid DNA in each colony was amplified in situ by tansferring the filters to fresh plates and incubating overnight at 37°. Finally, these replica filters were prepared for standard colony hybridization by treatment with base followed by neutralization and drying. The dried filters were baked in vacuo for 2 hours at 80° to fix the liberated DNA to the nitrocellulose.

Step F. Colony Hybridization

To identify EPA clones, one set of replica filters was probed with the set of 64 12-mers and the other corresponding set

of duplicate filters was probed with the 64 14-mers. The hybridizations were performed at room temperature in 4x SSC, 5x Denhardt's, 100 ug/ml Herring Sperm DNA and 10^6 ug/ml of ^{32}P-labelled oligonucleotide probe (labelled by polynucleotide kinase and ^{32}P-ATP). After overnight hybridization, the filters were washed with 2x SSC, 0.1% SDS at room temperature for 4-5 hours. Finally, the filters were exposed to film overnight at -80°C using intensifying screens.

Potential EPA clones were identified by aligning the duplicate filters and lookng for individual colonies which hybridized to both independent probes (the 64 14-mers and 64 12-mers). Ten such clones which appeared to hybridize to both probes were picked by aligning the x-ray film of the hybridized filter with the original master filter. These 10 colonies were picked and grown in liquid culture for the preparation of analytical amounts of DNA (prepared by the standard base treatment method).

Step G. Southern Analysis Of The Potential EPA Clones

Samples of DNA from each of the 10 clones were cleaved with PstI to excise the inserts. These digests wre analyzed by agarose gel electrophoresis. A variety of insert sizes were observed. The gel was next prepared for standard Southern analysis by incubating in base for 30 minutes, then in 1M Tris, 1.5M NaCl for 30 minutes to neutralize the base. Nitrocellulose strips were placed on either side of the gel. This sandwich was

placed on top of a 1 inch stack of paper towels and a second stack of towels was placed on top of the gel. DNA was allowed to leach out of the gel overnight (in both directions) giving replica filters of the gel. These filters were baked at 80°C for 2 hours in vacuo then probed overnight as described for the colony hybridization (1 filter with the 12 mers and 1 with 14 mers). After washing the filters and exposing them to film it was observed that the inserts from several clones hybridized to both probes but three clones (designated 27, 57, 58) hybridized most strongly to the probes. These three clones all clearly had at least one internal PstI site suggesting that they encoded the same sequence. To further test the clones, the Southern filters were washed to eliminate the radioactivity and were reprobed with the 2 pools of labeled 53-mers (hybridized as above except at 37°C). Both sets of 53-mers hybridized strongly only with the three clones designated 27, 57, and 58, again suggesting that the cDNA inserts of these clones were good candidates for the gene encoding EPA.

Step H. Sequencing The Candidate EPA Clones

To determine the sequence of the clones in the region expected to encode the amino acid sequence determined from the purified protein we first determined which of the 8 pools of eight 14-mers best hybridized to the clone. To do this each pool was labelled separately with polynucleotide kinase and ^{32}p-ATP and hybridized separately to a replica Southern filter prepared

as described above. Pool 14-5 proved to hybridize most strongly to clones 27, 57 and 58, indicating that this pool of oligonucleotides would be a good primer for dideoxy sequencing of the candidate clones.

For sequencing, CsCl purified clone 58 DNA was denatured with base and sequenced using standard dideoxy sequencing with pool 14-5 as primer. This primer gave an unambiguous DNA sequence which clearly encoded residues 19 back to 1. In addition clone 58 contained 17 codons from a potential signal peptide. From this result, we concluded that we had identified a clone encoding the purified EPA protein. Since we were confident that clones 27 and 57 also encoded the EPA polypeptide, we determined the complete nucleotide sequence of all three clones. This was accomplished by subcloning the PstI fragments of the clone (EPA tuned out to have 3 internal PstI sites which generate four small fragments for sequencing) into M13 and sequencing them using the universal M13 primer. To align the sequences of the four PstI fragments, olignonucleotide primers were synthesized for use in sequencing the plasmid DNA directly (as described above with pool 14-5). One of these primers;

d(T-G-C-A-C-C-T-G-T-G-T-C-C-C-A-C-C-))

encoded the end of the mature protein and could be used for insertion into the bacterial expression vector as well as sequencing at the 5' end of the coding region. A second primer;

(d(G-A-G-G-A-G-T-T-T-C-T-C-A-T-T-G-C-T-G))

encoded the region near the 3' end of the first PstI fragment and

could be used to sequence through the three PstI sites towards the 3' end of the clone. The final primer;

(d(C-C-A-C-A-A-G-C-A-A-T-G-A-G-G-T-G-C-C))

was complementary to the mRNA strand near the 5' end of the 3' Pst fragment of the clone and could be used to sequence towards the 5' end of the clone through the 3 Pst I sites. The complete DNA sequence of clone 57 is illustrated in Fig. 4 along with the deduced amino acid sequence of the translated EPA protein product and the coding region was the same for the other two clones.

Step I. Expression Of EPA In Mammalian Cells

To prove that the protein purified and cloned as EPA really is EPA it was necessary to express the cDNA clone and demonstrate biological activity. Examination of the DNA sequence yielded no convenient restriction sites to excise the EPA coding region (3 internal PstI sites made this enzyme useless). To get around this problem, clone 57 DNA was cleaved with Bgl I which cleaves in pBR322 127 nucleotides away from the unique Pst I site.

The Bgl I cut DNA was treated with the nuclease Bal 31 for a time sufficient to remove approximately 150 base pairs per end or 300 total from the DNA. The DNA was then cleaved with PvuI which cuts pBR322 126 nucleotides on the other side of the PstI site from the Bgl I site. This DNA was treated with the Klenow fragment of DNA polymerase I to render the ends blunt then ligated overnight with synthetic EcoRl linkers. The excess polymerized linkers were removed by digestion with EcoRl and the

approximately 800 base pair fragment was isolated from a preparative agarose gel. This DNA was recovered by the standard glass powder isolation method and was ligated to EcoRl linearized pBR322. Colonies containing EPA inserts were identified by colony hybridization using the labelled 14-mer (pool 14-5) as probe. Twelve colonies that scored as positive were picked, and their DNAs were prepared and analyzed by digestion with EcoRl. Three clones having inserts of the right size (about 800 base pairs) were sequenced using pool 14-5 as primer.

One of these clones, in which the EcoRl linker had been inserted into the 5' non-coding region of EPA, was used to prepare the insert to be expressed in animal cells. This 800 base pair EcoRl fragment was ligated with EcoRl digested transformation vector DNA. The ligation mix was transformed into E. coli and clones carrying the desired insert were identified by colony hybridization again, using labeled pool 14-5 DNA as probe. Twelve colonies which were identified as positive were used to prepare DNA. The orientation of their inserts was determined by digestion with AvaI, an enzyme which cleaves near the 3' end of the EPA coding region. One clone, which was determined to have the coding region oriented correctly in the vector was used to prepare DNA which was then purified by equilibrium density centrifugation in CsCl. This purified DNA (10ug) was used to transfect monkey COS-7 cells.

Step J. Construction Of Expression Vector p91023(B)

The vector pAdD26SVpA(3) was described by (Kaufman et al., Mol. Cell Biol. 2(11):1304-1319 [1982]. It has the structure illustrated in Fig. 1. Briefly this plasmid contains a mouse dihydrofolate reductase (DHFR) cDNA gene that is under transcriptional control of the adenovirus 2 (Ad2) major late promoter. A 5' splice site is included in the adenovirus DNA and a 3' splice site, derived from an immunoglobulin gene, is present between the Ad2 major late promoter and the DHFR coding sequence. The SV40 early polyadenylation site is present downstream from the DHFR coding sequence. The procaryotic-derived section of pAdD26SVpA(3) is from pSVOd (Mellon, P., Parker, V., Gluzman, Y. and Maniatis, T. 1981, Cell 27:279-288) and does not contain the pBR322 sequences known to inhibit replication in mammalian cells (Lusky, M., and Botchan, M. 1981, Nature (London) 293:79-81.

pAdD26SVpA(3) was converted into plasmid pCVSVL2 as illustrated in Fig. 1. pAdD26SVpA(3) was converted into plasmid pAdD26SVpA(3)(d) by deletion of one of the two PstI sites in pAdD26SVpA(3). This was accomplished by a partial digestion with PstI (using a deficiency of enzyme activity so that a subpopulation of linearized plasmids can be obtained in which only one PstI site is cleaved), then treatment with Klenow fragment of Pol I to blunt the ends, ligation to recircularize the plasmid, transformation of E. coli and screening for deletion of the PstI site located 3' of the SV40 polyadenylation sequence.

The adenovirus tripartite leader and virus associated genes (VA genes) were inserted into pAdD26SVpA(3)(d) as illustrated in Fig. 1. First, pAdD26SVpA(3)(d) was cleaved with PvuII to make a linear molecule opened within the 3' portion of the first of the three elements comprising the tripartite leader. Then, pJAW 43 (Zain et al. 1979, Cell 16 851) was digested with Xho I, treated with Klenow, digested with PvuII, and the 140 base pair fragment containing the second leader and part of the third leader was isolated by electrophoresis on an acrylamide gel (6% in Tris borate buffer; Maniatis et al. [1982] supra). The 140 bp fragment was then ligated to the PvuII digested pAdD26SVpA(3)(d). The ligation product was used to transform E. coli to tetracycline resistance and colonies were screened using the Grunstein-Hogness procedure using a ^{32}P labelled probe hybridizing to the 140 base pair fragment. DNA was prepared from positively hybridizing colonies to test whether the PvuII site reconstructed was 5'or 3' of the inserted 140 base pair DNA specific to the 2nd and 3rd adenovirus late leaders. In the correct orientation, the PvuII site is on the 5' side of the 140 base pair insert. This plasmid is designated pTPL in Fig. 1.

The Ava II D fragment of SV40 containing the SV40 enhancer sequence was obtained by digesting SV40 DNA with Ava II, blunting the ends with Klenow fragment of Pol I, ligating Xho I linkers to the fragments, digesting with Xho I to open the Xho I site, and isolating the fourth largest (D) fragment by gel electrophoresis. This fragment was then ligated to Xho I cut pTPL yielding the

plasmid pCVSVL2-TPL (Fig. 2). The orientation of the SV40 D Fragment in pCVSVL2-TPL was such that the SV40 late promoter was in the same orientation as the adenovirus major late promoter.

To introduce the adenovirus virus associated (VA) genes into the pCVSVL2-TPL, first a plasmid was constructed that contained the adenovirus type 2 Hind III B fragment. Adenovirus type 2 DNA was digested with Hind III and the B fragment was isolated after gel electrophoresis. This fragment was then inserted into pBR322 which had previously been digested with Hind III. After transformation of E. coli to ampicillin resistance, the recombinants were screened for insertion of the Hind III B fragment and the inserted orientation was determined by restriction enzyme digestion. pBR322 - Ad Hind III B contains the adenovirus type 2 Hind III B fragment in the orientation depicted in Fig. 2.

As illustrated in Fig. 2, the VA genes were conveniently obtained from plasmid pBR322-Ad Hind II by digesting with Hpa I, ligating EcoRl linkers and digesting with EcoRl, and recovering the 1.4kb fragment. This fragment having EcoRl sticky ends was then ligated into the EcoRl site of pCVSVL2-TPL (which had previously been digested with EcoRl). After transformation of E. coli HB101 and selection for tetracycline resistance, colonies were screened by filter hybridization to a DNA probe specific to the VA genes. DNA was prepared from positively hybridizing clones and characterized by restriction endonuclease digestion. The product plasmid was designated p91023.

The 2 EcoRl sites in p91023 were removed. p91023 was cut to completion with EcoRl, generating two DNA fragments: one, about 7Kb and, the other, a 1.4 Kb fragment containing the VA genes. The ends of both fragments are filled in using the Klenow fragment of PolI, and then both fragments, i.e. 1.3 Kb and 7Kb, were religated together. A plasmid p91023(A) containing the VA genes and similar to p91023 but deleted for the 2 EcoRl sites was identified by Grunstein-Hogness screening with the VA gene fragment, and by conventional restriction site analysis.

Then the single PstI site in p91023(A) was removed and replaced with an EcoRl site (Fig. 3). p91023(A) was cut to completion with PstI, and then treated with Klenow fragment of PolI to generate flush ends. EcoRl linkers were ligated to the blunted PstI site of p91023(A). The linear p91023(A), with EcoRl linkers attached at the blunted PstI site, was separated from unligated linkers and digested to completion with EcoRl, and then religated. A plasmid p91023(B) was recovered and identified to have a structure similar to p91023(A), but with an EcoRl site situated at the previous PstI site.

EPA clone 57 prepared as described above (Step I) was digested with EcoRl to obtain the DNA coding for EPA. This DNA was ligated into the EcoRl site of p91023(B). Monkey COS-7 cells were transfected with the EPA containing plasmids and grown to express EPA.

DNA transfections were performed as described by Sompayrac & Danna, Proc. Natl. Acad. Sci., 78, pp. 7575-8 (1981), with the

addition of chloroquin treatment (Luthman and Magnusson, Nuc. Acids Res. 11, pp. 1295-1308 (1983)). Cells were rinsed with serum free media, and incubated at 37 degrees for 12 hours in Dulbecco's modified Eagle's medium containing DEAE Dextran (MW 5,000,000 Pharmacia, 250 mg/ml) with .1M Tris pH 7.3 and 2 ug/ml of plasmid DNA. After incubation, the cells were rinsed with serum free media and treated 2 hours at 37 degrees with 1mM chlorquin in media containing serum. Cells were subsequently fed with media and incubated for indicated periods of time.

Conditioned medium from these cells greatly stimulated the growth of erythroid bursts and clusters in the standard EPA assay thereby demonstrating that the protein encoded by clone 57 is EPA.

Step K. Expression Of EPA In E. Coli

To express EPA in E. coli it was necessary to exactly fuse the mature protein coding region to the ATG found in the expression plasmid pALP-181. Clone 57 containing the EPA gene has been deposited in vector pALP-181 transformed in E. coli as ATCC No. 39823. To prepare a fragment useful for this, the EcoRl insert generated in Step I was cloned into the EcoRl site of M13mp9. Single stranded DNA was isolated from this phage preparation and used as a template with the primer repair oligonucleotide as primer. When priming DNA synthesis with the Klenow fragment of DNA polymerase, this primer generated a complementary strand that begins exactly with T-G-C-A-C-C-... the

sequence of the amino terminus of the mature protein. Single stranded regions of this molecule were removed by treatment with S1 nuclease and the coding region of EPA was excised by digestion with BamH1. The pALP-181 vector fragment was prepared by KpnI digestion which opens the plasmid immediately adjacent to the ATG to be used in expressing proteins. By treating with Klenow, the KpnI sticky ends are removed leaving a flush ATG in the desired location. Cleavage of the vector with BamH1 renders the vector capable of ligating with the flush to BamH1 fragment carrying the mature coding region of EPA. These two DNAs were ligated together and used to transform E. coli W3110 lambda Y139, a strain carrying a temperature sensitive lambda repressor protein (plating is done at 30° to keep the P_L promoter of pALP-181 turned off). Colonies from the transformation were lifted onto filters and two replica filters were prepared for colony hybridization. One replica was hybridized to the primer repair oligonucleotide and the second filter was hybridized to a junction oligonucleotide having the sequence d(T-A-C-A-T-A-T-G-T-G-C-A-C-C-T) which spans the desired correct sequence of the mature EPA sequence fused to the ATG of the expression plasmid. Four colonies which hybridized to both probes were grown and checked for the ability to produce EPA. To do this, 50 ml cultures were grown for several generations at 30°. To inactivate the repressor protein the temperature was shifted to 38° and incubation continued for two hours. The cells were harvested by centrifugation and lysed by sonication. The

extract was diluted in laemli SDS sample buffer and analyzed by SDS polyacrylamide gel electrophoresis. That the EPA polypeptide had been expressed was indicated by the appearance of a 28,000 dalton polypeptide in all four samples that was not seen in strains not carrying the EPA insert. That this protein was active was be demonstrated by assaying highly diluted samples of the extract. The EPA protein can be purified to substantial homogeneity by the procedure described in Step C above.

EXAMPLE 2

Step A: mRNA Preparation From The Gibbon T-Cells

A sample of the gibbon T-cell line designated UCD-MLA 144 was cultured for several weeks in RPMI 1640 (purchased from Gibco) and 20% fetal calf serum (FCS) until there was obtained 1 x 10^9 total cells. The cells were induced to produce high levels of EPA by activation for 24 hours in the presence of 10 nanograms per ml 12-0-tetradecanoylphorbol-13-acetate (TPA) in RPMI 1640 + 1% FCS. The cells were harvested by centrifugation (1000 rpm., 5 min.), washed once with phosphate buffered saline (PBS) and finally collected by centrifugation.

Membrane bound polysome (MBP) mRNA was prepared from these cells using the same procedure as described in Mechler, B. and Rabbitts, T.H. (1981), J. Cell Biol. 88:29.

Step B. First Strand cDNA Reaction

6 ug of MBP mRNA (from Step A) was diluted into a 50 ul cDNA

synthesis reaction mixture in accord with standard methods (Maniatis et al., supra) and the reaction initiated by the addition of reverse transcriptase. After incubation for 30 minutes at 42°C, the reaction was stopped by addition of EDTA to 50 mM, and diluted with H_2O to 100 ul. The mixture was extracted with phenol/chloroform and further extracted with chloroform. The cDNA/RNA hybrids were separated from unincorporated triphosphates by chromatography on a 2 ml Sepharose CL-4B column. The excluded fractions were pooled and the hybrids collected by ethanol precipitation. The final yield was 570 ng.

Step C. Second Strand cDNA Reaction

The first strand cDNA pellet (from Step B) was resuspended in 50 ml of H_2O, and second strand synthesis carried out in a standard reaction mixture with E. coli Polymerase I, E. coli ligase, and RNAse H. The reaction was incubated overnight at 16°C and then incubated for 1 hour at 37°C. The reaction was stopped by addition of EDTA and extracted with phenol/chloroform. The cDNA was separated from unincorporated triphosphates by chromatography on a Sepharose CL-4B column, the excluded fractions pooled and the cDNA collected by ethanol precipitation.

Step D. Recombinant cDNA Preparation

The cDNA pellet (from Step C) was resuspended in 75 ul of H_2O. Homopolymeric C "tails" were added to the ends of the cDNA by adding 10 ul of the cDNA solution to a 25 ul standard reaction

mixture with terminal transferase, and incubating at 30°C for 5 minutes. The reaction was stopped by the addition of EDTA to 40 mM and and heat inactivation at 68°C for 10 minutes. 10 ng of this tailed cDNA was annealed with 50 ng of G-tailed pBR322 (purchased from NEN) in 10 ul of 10 mM Tris, pH 7.5, 1 mM EDTA, and 100 mM NaCl. The annealing reaction was incubated for 10 minutes at 68°C and then for 2 hours at 57°.

Step E. Bacterial Transformation

E. coli strain MC1061, was grown in L-broth, chilled on ice, harvested by centrifugation, and treated wth $CaCl_2$ to prepare them for transformation. 5 ul of the cDNA annealing reaction was then incubated with 200 ul of the $CaCl^2$-treated bacteria. 15 such transformations were performed, using all of the annealed cDNA, and spread on 15 cm 1% agar L-broth plates containing 10 ug/ml tetracycline. Approximately 1000 colonies grew on each plate.

Step F. Replica Plating

10,000 colonies from the transformation were each picked with a toothpick, transferred to fresh plates (500 per plate in a grid), and grown overnight at 37°C. The colonies were then lifted from each plate by pressing a dry nitrocellulose filter firmly over the surface of the plate. Two replica filters were prepared from each of these master filters. The master filters were stored at 4°C, and the replica filters treated with base,

and baked to prepare them for hybridization.

Step G. Preparation of ^{32}P Labelled Hybridization Probes

The cDNA insert from EPA Clone 57 (Example 1) was isolated by digestion with the restriction enzyme EcoRl, and electrophoresis in an agarose gel with Tris acetate and ethidium bromide. The band containing the cDNA fragment was cut from the gel and purified by the glass powder technique.

300 ng of the cDNA fragment was then added to 1 ul of 10 x T4 DNA Polymerase Buffer (0.33 M Tris Acetate, pH 7.9, 0.66 M potassium acetate, 0.1 M Magnesium acetate and 10 mM dithiothreitol), and 3 units of T4 DNA Polymerase (New England Biolabs), and diluted with water to 10 ul. After incubation for 5-10 minutes at 37°C, this mixture was combined with 1 ul 10 x T4 DNA Polymerase Buffer; 1 ul of a 2 mM solution of each of dCTP, dTTP, dGTP; 10 ul of ^{32}PdATP (10 uCi/ul, 3,000 Ci/mmole); and 3 units of T4 DNA Polymerase. The reaction was incubated for 20 minutes at 37°C. Then 1 ul of 2 mM dATP was added and the reaction incubated for an additional 10 minutes at 37°C.

The unincorporated triphosphates were separated from the labelled cDNA by chromatography on a Sephadex G100 column. A second probe was prepared from a synthetic oligonucleotide having the sequence:

ATG GCC CCC TTT G

which is complimentary to the amino terminus of the EPA coding region. This oligonucleotide was labelled with ^{32}P dATP at its

5' end using a standard polynucleotide kinase reaction.

Step H. Isolation of Gibbon EPA cDNA Clones

In a standard hybridization screening procedure, some 45 clones hybridized with the T4 labelled EPA Clone 57 cDNA. Of these, approximately 12 also hybridized to the labelled oligonucleotide probe. Three of these clones were sequenced in accordance with conventional procedures (as described for example in Nature, Vol. 35, 27 June 1985:768-771) and each produced the EPA nucleotide gene sequence and deduced EPA amino acid protein sequence shown in Fig. 5.

The invention has been described in detail, including the preferred embodiments thereof. However, it will be appreciated that those skilled in the art, upon consideration of this specification, may make modifications and improvements within the spirit and scope of the invention.

What is claimed:

1. A method for preparing an expression vector containing an EPA gene, said method comprising:

inducing a T-lymphocyte cell to enhance lymphokine production;

preparing cytoplasmic RNA from said induced T-lymphocyte;

preparing polyadenylated messenger RNA from said cytoplasmic RNA;

preparing single stranded cDNA from said messenger RNA;

hybridizing said single stranded cDNA with a preparatior from B-cell lines or immature T-cell lines to form double stranded molecules;

separating the double stranded molecules from the remaining single stranded cDNA;

converting the single stranded cDNA to double stranded cDNA;

cloning the double stranded cDNA;

hybridizing the clones with a pool of oligonucleotide probes;

selecting the clone that hybridizes most strongly with the probes;

sequencing the clone that hybridizes most strongly to determine the oligonucleotide sequence of the EPA gene; and

inserting the EPA gene into an expression vector.

2. Human EPA protein substantially free of other protein of human origin.

3. Gibbon EPA protein substantially free of other protein of gibbon origin.

4. EPA protein substantially free of glycosylation.

5. EPA protein glycosylated by expression of a recombinant expression vector containing an EPA gene in a transformed eukaryotic cell.

6. cDNA that codes EPA.

7. An expression vector comprising cDNA coding for EPA.

8. A transformed cell comprising the expression vector of claim 7 or an allelic variation thereof.

9. A therapeutic composition for the treatment of mammals to stimulate growth and formation of erythroid cells comprising erythroid cell growth and formation stimulating treatment amount of EPA protein in a pharmacological carrier.

10. A method for the treatment of mammals having low red blood counts, said method comprising treating said mammal with EPA protein.

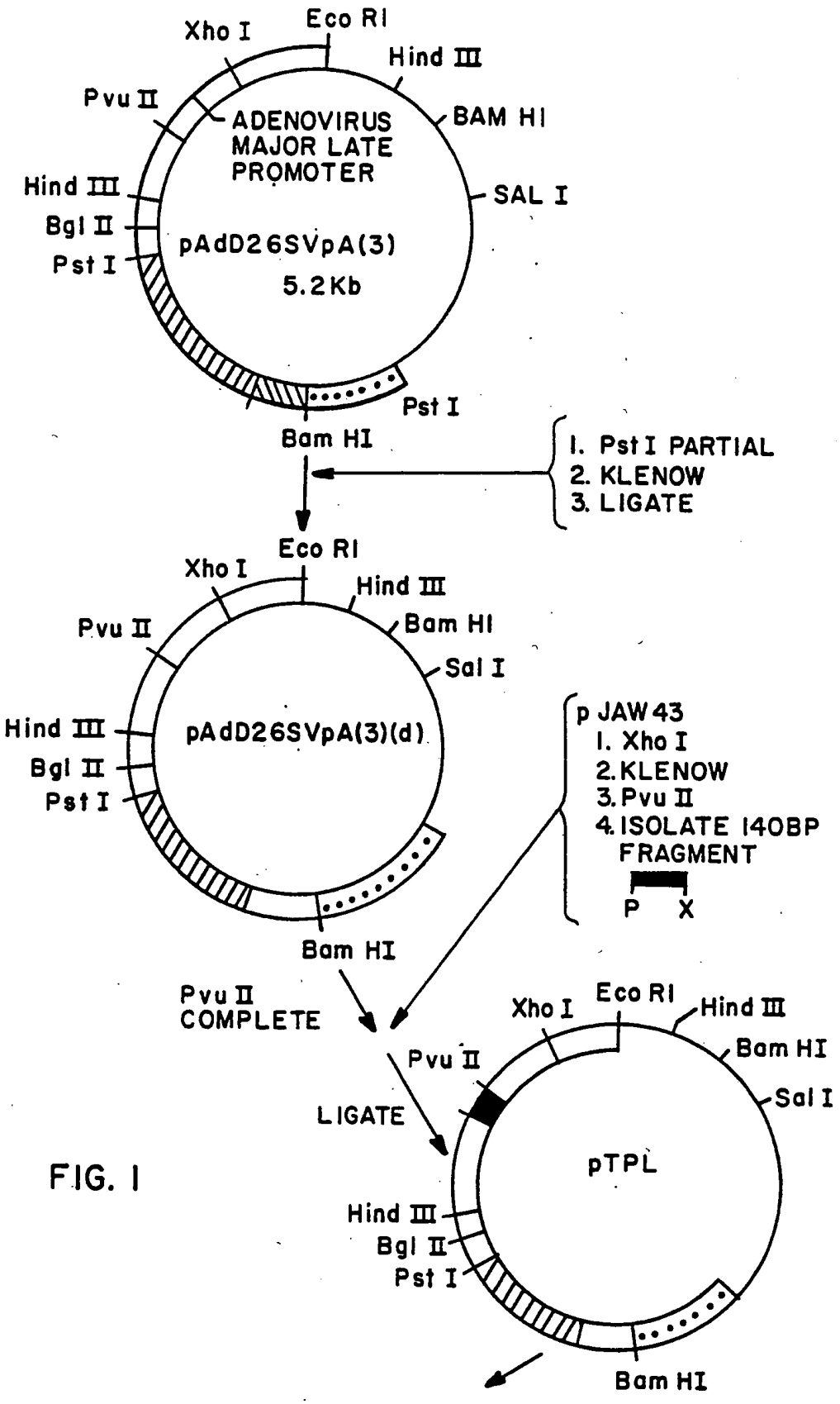
Xho I
Eco RI
Hind III
Pvu II
ADENOVIRUS
MAJOR LATE
PROMOTER
BAM HI
Hind III
SAL I
Bgl II
pAdD26SVpA(3)
Pst I
5.2Kb
Pst I
Bam HI
1. Pst I PARTIAL
2. KLENOW
3. LIGATE
Eco RI
Xho I
Hind III
Pvu II
Bam HI
Sal I
Hind III
pAdD26SVpA(3)(d)
Bgl II
Pst I
p JAW 43
1. Xho I
2. KLENOW
3. Pvu II
4. ISOLATE 140BP
FRAGMENT
P X
Bam HI
Pvu II
COMPLETE
Xho I
Eco RI
Hind III
Bam HI
Pvu II
Sal I
LIGATE
pTPL
Hind III
Bgl II
Pst I
Bam HI

FIG. 1

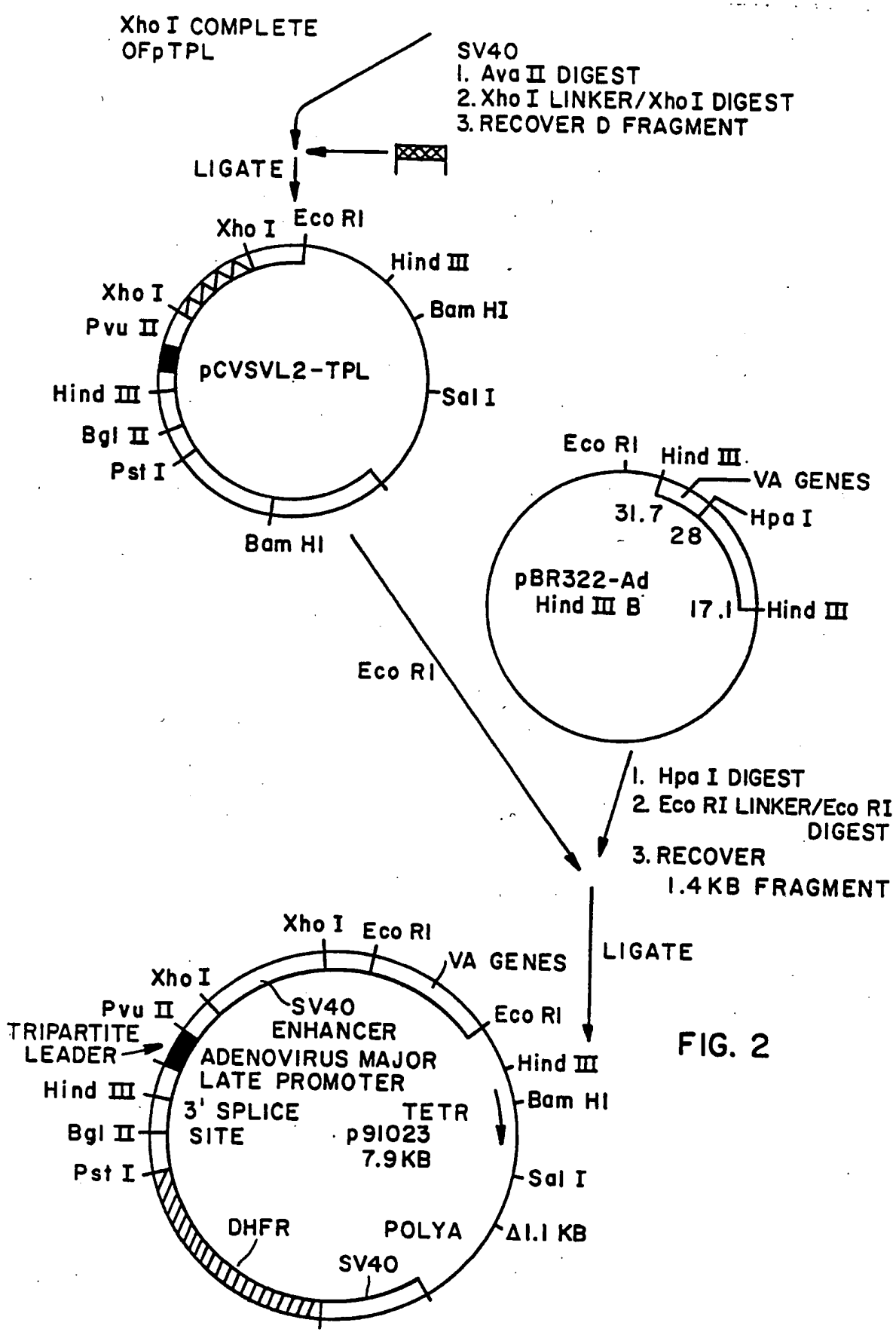
Xho I COMPLETE
OFpTPL
SV40
1. Ava II DIGEST
2. Xho I LINKER/Xho I DIGEST
3. RECOVER D FRAGMENT
LIGATE
Xho I
Eco RI
Hind III
Bam HI
Sal I
Pst I
Bgl II
Pvu II
pCVSVL2-TPL
Bam HI
Eco RI
Hind III
VA GENES
Hpa I
31.7
28
pBR322-Ad
Hind III B
17.1
Hind III
Eco RI
1. Hpa I DIGEST
2. Eco RI LINKER/Eco RI DIGEST
3. RECOVER
1.4 KB FRAGMENT
LIGATE
Xho I
Eco RI
VA GENES
Xho I
Pvu II
TRIPARTITE
LEADER
SV40
ENHANCER
ADENOVIRUS MAJOR
LATE PROMOTER
3' SPLICE
SITE
TETR
p91023
7.9KB
Eco RI
Hind III
Bam HI
Sal I
Hind III
Bgl II
Pst I
DHFR
POLYA
Δ1.1 KB
SV40

FIG. 2

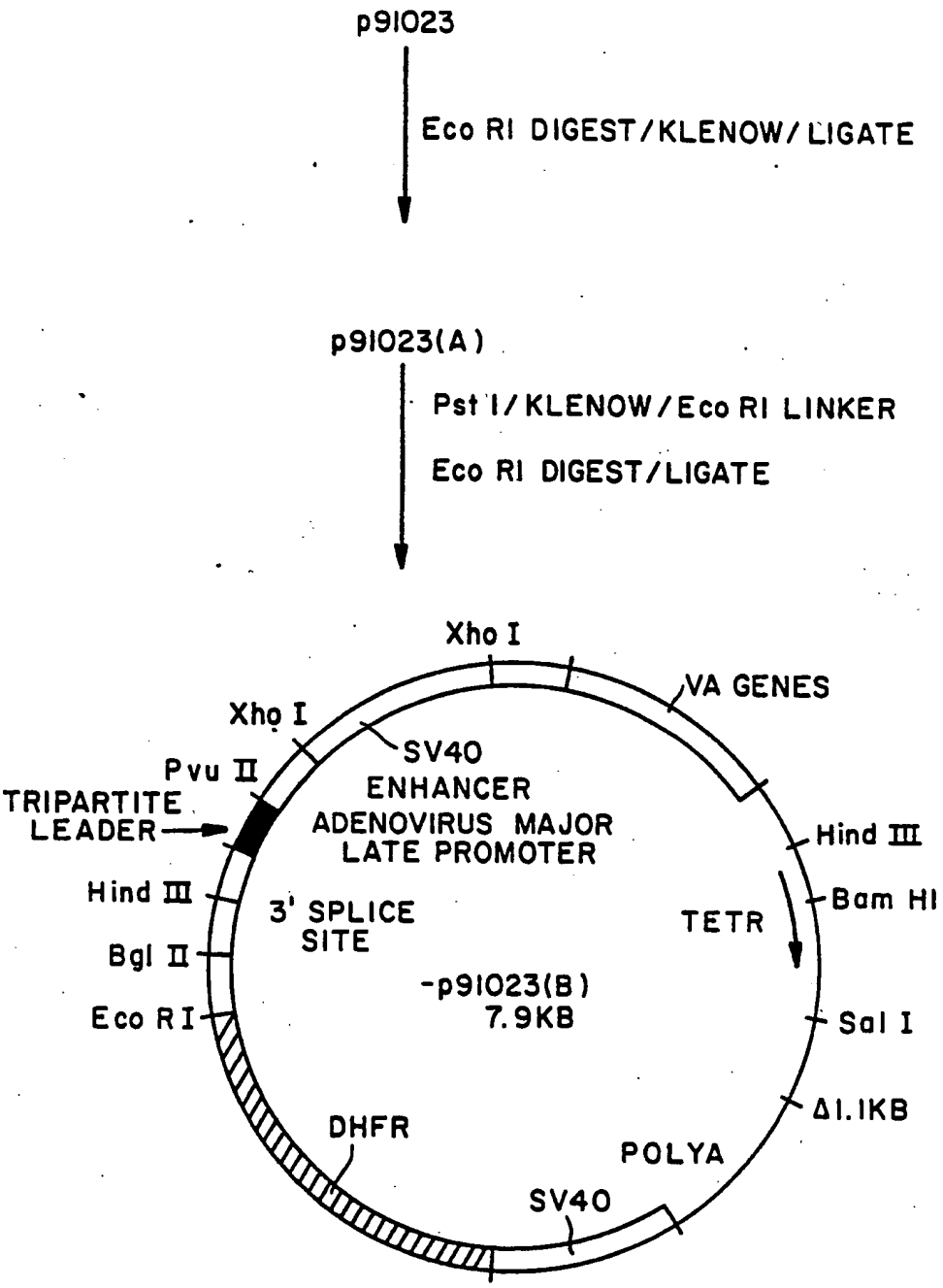
p91023
Eco RI DIGEST/KLENOW/LIGATE
p91023(A)
Pst I/KLENOW/Eco RI LINKER
Eco RI DIGEST/LIGATE
Xho I
VA GENES
Xho I
SV40
ENHANCER
ADENOVIRUS MAJOR
LATE PROMOTER
Pvu II
TRIPARTITE
LEADER
Hind III
3' SPLICE
SITE
Bgl II
Eco RI
-p91023(B)
7.9KB
Hind III
Bam HI
TETR
Sal I
Δ1.1KB
DHFR
POLYA
SV40

FIG. 3

```
          10          20          30          40
CCCGCAGATC CAGCGCCCAG AGAGACACCA GAGAACCCAC C ATG GCC CCC TTT
                                                MET Ala Pro Phe

 56                  71                  86
GAG CCC CTG GCT TCT GGC ATC CTG TTG TTG CTG TGG CTG ATA GCC
Glu Pro Leu Ala Ser Gly Ile Leu Leu Leu Leu Trp Leu Ile Ala

101                 116                 131
CCC AGC AGG GCC TGC ACG TGT GTC CCA CCC CAC CCA CAG ACG GCC
Pro Ser Arg Ala Cys Thr Cys Val Pro Pro His Pro Gln Thr Ala

146                 161                 176
TTC TGC AAT TCC GAC CTC GTC ATC AGG GCC AAG TTC GTG GGG ACA
Phe Cys Asn Ser Asp Leu Val Ile Arg Ala Lys Phe Val Gly Thr

191                 206                 221
CCA GAA GTC AAC CAG ACC ACC TTA TAC CAG CGT TAT GAG ATC AAG
Pro Glu Val Asn Gln Thr Thr Leu Tyr Gln Arg Tyr Glu Ile Lys

236                 251                 266
ATG ACC AAG ATG TAT AAA GGG TTC CAA GCC TTA GGG GAT GCC GCT
MET Thr Lys MET Tyr Lys Gly Phe Gln Ala Leu Gly Asp Ala Ala

281                 296                 311
GAC ATC CGG TTC GTC TAC ACC CCC GCC ATG GAG AGT GTC TGC GGA
Asp Ile Arg Phe Val Tyr Thr Pro Ala MET Glu Ser Val Cys Gly

326                 341                 356
TAC TTC CAC AGG TCC CAC AAC CGG AGC GAG GAG TTT CTC ATT GCT
Tyr Phe His Arg Ser His Asn Arg Ser Glu Glu Phe Leu Ile Ala

371                 386                 401
GGA AAA CTG CAG GAT GGA CTC TTG CAC ATC ACT ACC TGC AGT TTT
Gly Lys Leu Gln Asp Gly Leu Leu His Ile Thr Thr Cys Ser Phe

416                 431                 446
GTG GCT CCC TGG AAC AGC CTG AGC TTA GCT CAG CGC CGG GGC TTC
Val Ala Pro Trp Asn Ser Leu Ser Leu Ala Gln Arg Arg Gly Phe

461                 476                 491
ACC AAG ACC TAC ACT GTT GGC TGT GAG GAA TGC ACA GTG TTT CCC
Thr Lys Thr Tyr Thr Val Gly Cys Glu Glu Csy Thr Val Phe Pro

506                 521                 536
TGT TTA TCC ATC CCC TGC AAA CTG CAG AGT GGC ACT CAT TGC TTG
Cys Leu Ser Ile Pro Cys Lys Leu Gln Ser Gly Thr His Cys Leu

551                 566                 581
TGG ACG GAC CAG CTC CTC CAA GGC TCT GAA AAG GGC TTC CAG TCC
Trp Thr Asp Gln Leu Leu Gln Gly Ser Glu Lys Gly Phe Glu Ser

596                 611                 626
CGT CAC CTT GCC TGC CTG CCT CGG GAG CCA GGG CTG TGC ACC TGG
Arg His Leu Ala Cys Leu Pro Arg Glu Pro Gly Leu Cys Thr Trp

641                 656                 675         685
CAG TCC CTG CGG TCC CAG ATA GCC TGA ATCCTGCCCG GAGTGGAAGC
Gln Ser Leu Arg Ser Gln Ile Ala

        695         705         715         725         735
TGAAGCCTGC ACAGTGTCCA CCCTGTTCCC ACTCCCATCT TTCTTCCGGA

        745         755         765
CAATGAAATA AAGAGTTACA CCAGCAAAAA AAAAA
```

FIG. 4

```
      6          16                                          42
GCCCAG AGAGACACCA GAGAACCCAC C ATG GCC CCC TTT GAG CCC
                               MET Ala Pro Phe Glu Pro

                                                          87
CTG GCT TCT GGC ATC CTG TTG TTG CTG TGG CTG ATA GCC CCC AGC
Leu Ala Ser Gly Ile Leu Leu Leu Leu Trp Leu Ile Ala Pro Ser

AGG GCC TGC ACC TGT GTC CCA CCC CAC CCA CAG ACG GCC TTC TGC
Arg Ala Cys Thr Cys Val Pro Pro His Pro Glh Thr Ala Phe Cys

                  150
AAT TCC GAC CTC GTC ATC AAG GCC AAG TTT GTG GGG ACA CCA GAA
Asn Ser Asp Leu Val Ile Lys Ala Lys Phe Val Gly Thr Pro Glu

              192
GTC AAC CAG ACC ACC TTA TAC CAG CGT TAT GAG ATC AAG ATG ACC
Val Asn Gln Thr Thr Leu Tyr Gln Arg Tyr Glu Ile Lys MET Thr

                  240
AAG ATG TAC AAA GGG TTC CAA GCC TTA GGG GAT GCC GCT GAC ATC
Lys MET Tyr Lys Gly Phe Gln Ala Leu Gly Asp Ala Ala Asp Ile

                                  300
CGG TTC GTC TAC ACC CCC GCC ATG GAG AGT GTC TGC GGA TAC TTC
Arg Phe Val Tyr Thr Pro Ala Met Glu Ser Leu Val Gly Tyr Phe

                                                      360
CAC AGG TCC CAC AAC CGC AGC GAG GAG TTT CTC ATT GCT GGA AAA
His Arg Ser His Asn Arg Ser Glu Glu Phe Leu Ile Ala Gly Lys

CTG CAG GAC GGA CTC TTG CAC ATC ACC ACC TGC AGT TTC GTG GCT
Leu Gln Asp Gly Leu Leu His Ile Thr Thr Gys Ser Phe Val Ala

                  420                                     450
CCC TGG AAC AGT CTG AGC TTA GCT CAG CGC CGG GGC TTC ACC AAG
Pro Trp Asn Ser Leu Ser Leu Ala Gln Arg Arg Gly Phe Thr Lys

ACC TAC ACT GTT GGC TGT GAG GAA TGC ACA GTG TTT CCC TGT TTA
Thr Tyr Thr Val Gly Cys Glu Glu Csy Thr Val Phe Pro Cys Leu

                  510
TCC ATC CCC TGC AAA CTG CAG AGT GGC ACT CAT TGC TTG TGG ACG
Ser Ile Pro Cys Lys Leu Gln Ser Gly Thr His Cys Leu Trp Thr

                                  570
GAC CAG CTC CTC CAA GGC TCT GAA AAG GGC TTC CAG TCC CGT CAC
Asp Gln Leu Leu Gln Gly Ser Glu Lys Gly Phe Gln Ser Arg His

                                                      630
CTT GCC TGC CTG CCT CGG GAG CCA GGG CTG TGC ACC TGG CAG TCC
Leu Ala Cys Leu Pro Arg Glu Pro Gly Leu Cys Thr Trp Gln Ser

                                                671
CTG CGG TCC CAG ATA GCC TGA ATCCTGCCCG GAGTGGAAGC
Leu Arg Ser Gln Ile Ala

                                  701          711
TGAAGCCTGC ACAGTGTCCA CCCTGTTCCC ACTCCCATCT TTCCTCTGGA

      731         741
CAATGAAATA AAGAGTTACC ACCCAGCG
```

FIG. 5

SUBST

INTERNATIONAL SEARCH REPORT

International Application No PCT/US85/01900

I. CLASSIFICATION OF SUBJECT MATTER (If several classification symbols apply, indicate all) [3]

According to International Patent Classification (IPC) or to both National Classification and IPC

Int. Cl.4 C12P 21/00, C12N 15/00, C12N 5/00, C07H 15/12, A23J 1/00, A61K 39/40

II. FIELDS SEARCHED

Minimum Documentation Searched [4]

Classification System	Classification Symbols
U.S.	435/68, 172.3, 240, 317; 935/11, 13, 32, 70, 71; 536/27; 260/112R, 112.5R

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched [5]

COMPUTER DATA BASES: CHEM. ABSTRACTS FILE 309, 310, 320, 311
BIOSIS FILE 5, 55, 255

III. DOCUMENTS CONSIDERED TO BE RELEVANT [14]

Category *	Citation of Document, [16] with indication, where appropriate, of the relevant passages [17]	Relevant to Claim No. [18]
X, Y	US, A, 4,438,032, Published 20 March 1984	1,2,6-8 3-5,9,10
Y	N, Taniguchi et al, Nature Vol. 302 pp 305-310 24 March 1983	3-5
X, Y	N, Koeffler et al, Blood Vol. 64 pp 482-490 August 1984	2, 9,10

* Special categories of cited documents: [15]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search [2]	Date of Mailing of this International Search Report [2]
12/17/85	27 DEC 1985
International Searching Authority [1]	**Signature of Authorized Officer [20]**
ISA/US	Alvin E. Tanenholtz

FURTHER INFORMATION CONTINUED FROM THE SECOND SHEET

V. ☐ OBSERVATIONS WHERE CERTAIN CLAIMS WERE FOUND UNSEARCHABLE [10]

This International search report has not been established in respect of certain claims under Article 17(2) (a) for the following reasons:

1. ☐ Claim numbers ....., because they relate to subject matter [12] not required to be searched by this Authority, namely:

2. ☐ Claim numbers ....., because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out [13], specifically:

VI. ☒ OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING [11]

This International Searching Authority found multiple inventions in this international application as follows:

I Claims 1 and 6-8

II Claims 2-5, 9 and 10

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims of the international application. Telephone Interview.

2. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims of the international application for which fees were paid, specifically claims:

3. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claim numbers:

4. ☐ As all searchable claims could be searched without effort justifying an additional fee, the International Searching Authority did not invite payment of any additional fee.

Remark on Protest

☐ The additional search fees were accompanied by applicant's protest.

☐ No protest accompanied the payment of additional search fees.